September 2019 · Kate Saslow and Philippe Lorenz

Artificial Intelligence Needs Human Rights

How the focus on ethical AI fails to address privacy, discrimination and other concerns

Stiftung Neue Verantwortung

Think Tank at the intersection of technology and society

Executive Summary

AI has been a catalyst for automation and efficiency in numerous ways, but has also had harmful consequences, including: unforeseen algorithmic bias that affects already marginalized communities, as with Amazon's AI recruiting algorithm that showed bias against women; accountability and liability coming into question if an autonomous vehicle injures or kills, as seen with Uber's self-driving car casualties; even the notion of democracy is being challenged as the technology enables authoritarian and democratic states like China and the United States to practice surveillance at an unprecedented scale.

The risks as well as the need for some form of basic rules have not gone unnoticed and governments, tech companies, research consortiums or advocacy groups have broached the issue. In fact, this has been the topic of local, national, and supranational discussion for some years now, as can be seen with new legislation popping up to ban facial recognition software in public spaces. The problem with these discussions, however, is that they have been heavily dominated by how we can make AI more "ethical". Companies, states, and even international organizations discuss ethical principles, such as fair, accountable, responsible, or safe AI in numerous expert groups or ad hoc committees, such as the High-Level Expert Group on AI in the European Commission, the group on AI in Society of the Organization for Economic Co-operation and Development (OECD), or the select committee on Artificial Intelligence of the United Kingdom House of Lords.

This may sound like a solid approach to tackling the dangers that AI poses, but to actually be impactful, these discussions must be grounded in rhetoric that is focused and actionable. Not only may the principles be defined differently depending on the stakeholders, but there are overwhelming differences in how principles are interpreted and what requirements are necessary for them to materialize. In addition, ethical debates on AI are often dominated by American or Chinese companies, which are both propagating their own idea of ethical AI, but which may in many cases stand in conflict with the values of other cultures and nations. Not only do different countries have different ideas of which "ethics" principles need to be protected, but different countries play starkly different roles in developing AI. Another problem is when ethical guidelines are discussed, suggestions often come from tech

companies themselves, while voices from citizens or even governments are marginalized.

Self-regulation around ethical principles is too weak to address the spreading implications that AI technologies have had. Ethical principles lack clarity and enforcement capabilities. We must stop focusing the discourse on ethical principles, and instead shift the debate to human rights. Debates must be louder at the supranational level. International pressure must be put on states and companies who fail to protect individuals by propagating AI technologies that carry risks. Leadership must be defined not by actors who come up with new iterations of ethical guidelines, but by those who develop legal obligations regarding AI, which are anchored in and derived from a human rights perspective.

A way to do this would be to reaffirm the human-centric nature of AI development and deployment that follows actionable standards of human rights law. The human rights legal framework has been around for decades and has been instrumental in fighting and pressuring states to change domestic laws. Nelson Mandela referred to the duties spelled out in the Universal Declaration of Human Rights while fighting to end apartheid in South Africa; in 1973 with *Roe v. Wade* the United States Supreme Court followed a larger global trend of recognizing women's human rights by protecting individuals from undue governmental interference in private affairs and giving women the ability to participate fully and equally in society; more recently, open access to the Internet has been recognized as a human right essential to not only freedom of opinion, expression, association, and assembly, but also instrumental in mobilizing the population to call for equality, justice, and accountability in order to advance global respect for human rights. These examples show how human rights standards have been applied to a diverse set of domestic and international rules. That these standards are actionable and enforceable show that they are well-suited to regulate the cross-border nature of AI technologies. AI systems must be scrutinized through a human rights perspective to analyze current and future harms either created or exacerbated by AI, and take action to avoid any harm.

The adoption of AI technologies has spread across borders and has had diverse effects on societies all over the world. A globalized technology needs international obligations to mitigate the societal problems being faced at an accelerated and larger scale. Companies and states should strive for the development of AI technologies that uphold human rights. Centering the AI discourse around human rights rather than simply ethics can be one way of providing a clearer legal basis for development and deployment of AI techno-

logies. The international community must raise awareness, build consensus, and analyze thoroughly how AI technologies violate human rights in different contexts and develop paths for effective legal remedies. Focusing the discourse on human rights rather than ethical principles can provide more accountability measures, more obligations for state and private actors, and can redirect the debate to rely on consistent and widely accepted legal principles developed over decades.

Table of Contents

Introduction 6

Chapter 1: Problem Analysis 7
Ethical implications of AI development in the private sector 7
Ethical implications of how AI is used by governments 9

Chapter 2: A Discourse Focused on Ethical AI 11

Chapter 3: Need to Tie the Ethical Debate to Human Rights 14
The human rights framework provides legal basis 15
AI can build on digital rights' legal precedents 17
Globalized technologies require international obligations 19

Conclusion 21

SNV's Artificial Intelligence and Foreign Policy project was made possible by the generous support from the German Federal Foreign Office and the Mercator Foundation. The views expressed in this policy brief do not necessarily represent the official positions of the German Federal Foreign Office or the Mercator Foundation.

The authors would like to thank Mallory Knodel, Vidushi Marda, and Jonathan Andrew for their inspiration and valuable feedback during the writing of this policy brief.

Introduction

Artificial intelligence (AI)[1] has promised to revolutionize daily tasks, supercharge industry, gather and make sense of citizen data for good governance and data-driven policy making, and more. Harnessing the potential of AI has been pursued by individuals, corporations, and governments. Policy makers have been pulled in many directions, including how to encourage domestic industry to jump on the bandwagon and invest in AI research and development (R&D) while simultaneously researching ethical AI and using AI for good. Given the fast nature of AI development in recent years, fulfilling both goals has proven tricky.

There appears to be a dissonance between the economic incentives of developing AI products and swiftly placing them onto the market on the one hand, and promoting consumer protection and ensuring innovators take time to develop certain, safe, responsible products with ethical outcomes on the other hand. Governments, companies, and other stakeholders are aware of these risks, and there has been a growing discourse around ethical guidelines and principles to inform AI design and use. But there are notable gaps in this discourse; documents use broad and lofty language lacking clarity in terms of what they mean in practice. They also lack accountability mechanisms. Adding to the ethics discourse by anchoring it in the human rights legal framework is a way to demand actors be held accountable for any consequences of AI products that infringe upon universal human rights.

This paper highlights first how the increasing use of AI technologies in the private and public sectors has the potential to cause unethical outcomes.

1 AI traditionally refers to the scientific pursuit of teaching machines to think like humans (Francois Chollet, *Deep Learning with Python* (Shelter Island: Manning, 2018)). AI is often used as an umbrella term that covers several subdisciplines. Machine learning is one common subfield of AI, in which a machine *learns* from training data by automatically extracting patterns or features, and is able to make *inferences* following this feature extraction when presented with new data. (Philippe Lorenz & Kate Saslow, "Demystifying AI & AI Companies, Stiftung Neue Verantwortung, July 2019, https://www.stiftung-nv.de/sites/default/files/demystifying_ai_and_ai_companies.pdf). When referring to AI in this paper, we simply address the results and decisions that systems or software reach by machine learning technologies, with which consumers and citizens interact on a daily basis. There are other important subfields of AI that are not machine learning. For a historic overview of the two divergent paths of AI research (Symbolic versus Connectionist), see: Dominique Cardon, Jean-Philippe Cointet, and Antoine Mazieres, "Neurons Spike Back: The invention of inductive machines and the artificial intelligence controversy", *Reseaux* 2018/5 vol. 211, pp. 173-220, https://neurovenge.antonomase.fr/NeuronsSpikeBack.pdf.

Next, it shows how the discourse surrounding AI has changed over time. As stakeholders became aware of ethical concerns, they started to engage more in a discourse that fostered ethical guidelines, principles and values that were supposed to make consumers feel safer about AI deployment. The last chapter discusses how anchoring the ethics discussion in the international human rights legal framework is one way to address specific shortcomings of the ethics debate and make the international discourse stronger, clearer, and more capable of governing AI technologies.

Chapter 1: Problem Analysis

We live in a world with increasing amounts of data. With the rapid spread of the Internet and connected devices, users produce data constantly through interaction with digital services, and machines steadily collect and process data through sensors. With the growth of Big Data came the development of tools to analyze and process these data. Advancements in machine learning (ML), deep learning (DL), and other AI technologies have shown huge potential for efficiency gains from automation, and improvements in data analysis and statistical inference. Private sector actors quickly realized that better predictions and decisions could be made, which changed business models of traditional products and services. Public sector actors followed in this logic and began to implement AI systems to gather and analyze citizen data in order to provide better services in agencies overwhelmed by lack of public servants and resources. There is a general hype behind AI and its potential to help both businesses and governments. But on both levels, the promises of AI unleash serious potential for bias, discrimination, and harm.

Ethical implications of AI development in the private sector

The ubiquitous role of AI in our day to day lives – both in public and private spaces – could infringe upon the human rights of citizens all over the world. The scale and impact of these infringements are not exactly clear. Additionally, AI development in the private sector has followed the quick and dirty paradigm that Facebook popularized: "move fast and break things". There is a discrepancy between AI market dynamics and development of products that respect the basic rights of individuals, such as right to privacy, the freedom from discrimination, freedom of expression, and more. The market favors innovation and speed, and may overlook qualities like accuracy or certainty. Stakeholders from advocacy groups, research institutions, academia,

government, and even employees of Big Tech[2] have come forward and called for this paradigm to change.

Harmful AI stemming from the design of a system can occur, for example, from poor quality or biased data, algorithmic bias, or opaque models. While the tool may have been a product of good intentions, lack of foresight and impact analyses led to harmful technologies running rampant. For example when credit underwriting agencies realized that people with little to no credit had difficulty obtaining loans, some Fintech[3] (financial technology) startups had the idea to develop tools to gather more information on subjects simultaneously and analyze this Big Data using algorithms to help clients obtain loans based on data other than existing credit. Available data sometimes came from social media and other personal and often sensitive sources. Using other data points led to discriminatory outcomes – for example, proxies such as which high school a subject attended could be used to infer zip code, which is often criticized as being "correlated with a predictive characteristic that is left out of the model (e.g. race)", and may ultimately lead to racist credit underwriting.[4] Although the hope was to remove potential for human biases from the credit underwriting equation, relying on biased algorithms or proxy data creates its own harm.

Another ethical implication of AI development in the private sector is when products are rolled out and sold before they ensure a certain level of accuracy or certainty. An example of this can be seen with Amazon's facial recognition software, Rekognition. This technology caused controversy[5] among civil liberties advocacy groups, as the ACLU (American Civil Liberties Union) discovered that the software falsely matched faces of color (and especially

2 James Vincent, "Google employee who helped lead protests leaves company", *The Verge*, 16 July 2019: https://www.theverge.com/2019/7/16/20695964/google-protest-leader-meredith-whittaker-leaves-company.

3 For an overview of AI applications to Fintech in addition to credit underwriting, see: Corbin Hudson, "Ten Applications of AI to Fintech", *Towards Data Science*, 28 November 2018, https://towardsdatascience.com/ten-applications-of-ai-to-fintech-22d626c2fdac.

4 Aaron Klein, "Credit denial in the age of AI", *Brookings,* 11 April 2019, https://www.brookings.edu/research/credit-denial-in-the-age-of-ai/.

5 It is important to note that controversy around Rekognition was sparked for reasons other than just accuracy or certainty; namely, the fundamental discord between civil liberties and the government procurement and application of facial recognition in the public sphere.

women's and darker-skinned faces[6]) disproportionately with faces recorded in a database of arrest photos.[7] Rekognition highlights problematic market dynamics that lead to unethical outcomes: because Amazon invested heavily into the development of this software, they expect a return on their investment. This can mean pushing products onto the market despite substandard accuracy. That the main buyers of Rekognition are currently governmental law enforcement and immigration agencies in the United States can lead to further rights violations. In communities where people of color may already be subject to harmful police practices (e.g. near the US-Mexico border), being falsely "identified" by Rekognition can cost individuals their freedom and livelihood.

Ethical implications of how AI is used by governments

The use of AI technologies by governments can similarly lead to harmful outcomes. Public sector deployment of AI may perpetuate biases from technology developed in the private sector, particularly where many public authorities are not sufficiently resourced or under budget constraints and turn to AI to automate specific processes. It is understandable that government agencies want to take advantage of efficiency gains that automation through AI promises. However, questions of accuracy and biased models comes into play here, as well as in the private sector. Government procurement of AI technologies may further propagate the harmful effects that AI technologies can have on citizens.

The German Federal Office for Migration and Refugees (BAMF) has recently been scrutinized for their use of dialect-recognition software to determine asylum seekers' countries of origin during review of asylum applications. The BAMF's motivation to use dialect-recognition technologies comes from reasonable concerns: the majority of asylum seekers don't have identification

6 James Vincent, "Gender and racial bias found in Amazon's facial recognition technology (again)", *The Verge*, 25 January 2019, https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bias-race-gender.

7 Jacob Snow, "Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots, American Civil Liberties Union, 26 July 2018, https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28.

papers[8] upon their arrival in Germany, and the federal agency simply cannot hire enough staff, especially those with specific foreign language skills, to serve the influx of asylum seekers. Even in cases where asylum seekers have valid travel documents, dialect recognition may be used to confirm aspects of the asylum claim. Automating the asylum review process with dialect-recognition software could, in theory, benefit staffers and migrants alike. The problem in practice, however, was the accuracy of the software, and the reliability of the training data.[9] Linguists and computer scientists both criticized the use of the software before it was put into practice.[10] As of April 2018, however, the software had been used in just under 10,000 cases in Germany alone.[11] Similar to the use of Rekognition by the US immigrations agency, relying on software to help make decisions that significantly affect the livelihood[12] and liberties of already marginalized communities presents serious risks.

Government use of AI tools to exploit efficiency gains can also be witnessed in China – but on a completely different scale. China has been using machine learning (ML) powers to supercharge surveillance of the population and crack down and control the Uyghur minority.[13] It is reportedly not uncommon for the Chinese police to summon citizens of the ethnic minority to the police station to have their faces scanned from 360 degrees, give blood and DNA samples, and even be forced to speak for one minute to secure a voice

8 *Die Welt*, "Wie viele Flüchtlinge kommen ohne Papiere? Das BAMF weiß es nicht", 5 January 2017, https://www.welt.de/politik/deutschland/article160902357/Wie-viele-Fluechtlinge-kommen-ohne-Papiere-Das-BAMF-weiss-es-nicht.html (in German).

9 In this context specifically, reliability of training data comes into question because, for example, building a dataset from recordings of people from certain areas may not include all local dialects or accents spoken in that region.

10 Amar Toor, "Germany to use voice analysis software to help determine where refugees come from", *The Verge*, 17 March 2017, https://www.theverge.com/2017/3/17/14956532/germany-refugee-voice-analysis-dialect-speech-software.

11 Official answer of the German Government, "Einsatz von Spracherkennungssoftware durch das Bundesamt für Migration und Flüchtlinge", 16 April 2018, http://dipbt.bundestag.de/dip21/btd/19/016/1901663.pdf (in German).

12 In this context, livelihood may mean the difference between an asylum claim being accepted or denied.

13 Human Rights Watch, "How Mass Surveillance Works in Xinjiang, China: 'Reverse Engineering' Police App Reveals Profiling and Monitoring Strategies", 2 May 2019, https://www.hrw.org/video-photos/interactive/2019/05/02/china-how-mass-surveillance-works-xinjiang.

sample. Chinese AI tech company iFlytek works closely with law enforcement agencies to provide voiceprint technology.[14] Human Rights Watch even suspects that the company is in collaboration with the Chinese Ministry of Public Security to pilot a surveillance system that monitors telephone conversations.[15] The supercharged surveillance abilities that ML provides to private and state actors is a huge concern of our time.

Civil society advocacy groups, academics and researchers, private actors, governments and non-governmental agencies alike have begun to scrutinize the way AI products can lead to violations of human rights. But what are effective ways to address these problems? So far we have seen a growing focus on the development of ethical guidelines and principles of AI to guard against potential abuses.

Chapter 2: A Discourse Focused on Ethical AI

As it stands today, the most powerful players in the artificial intelligence arena are private sector actors. Innovation in AI comes from large companies and enterprises who possess state-of-the-art hardware, star researchers, and proprietary data to train and develop powerful AI models – whether to identify and classify images or whether to defeat the best humans in games like Go or poker.[16] Big Tech companies at the forefront of AI have become household names: the so-called GAFAM group of US businesses (Google, Amazon, Facebook, Apple, and Microsoft), and Baidu, Alibaba and Tencent in China. Because the dominant actors are corporate – not only developing the AI themselves, but bringing the AI to the market and potentially affecting billions of users – it is logical that solutions to this ethics dilemma have been demanded at the corporate level. The "faster is better" paradigm that drives corporate AI innovation must change. Governments and companies are aware of these risks, thus they are engaging in a discourse around ethics

14 Human Rights Watch, "China: Voice Biometric Collection Threatens Privacy", 22 October 2017, https://www.hrw.org/news/2017/10/22/china-voice-biometric-collection-threatens-privacy.

15 Isobel Cockerell, "Inside China's Massive Surveillance Operation", *Wired,* 9 May 2019, https://www.wired.com/story/inside-chinas-massive-surveillance-operation/.

16 Noam Brown and Tuomas Sandholm, "Superhuman AI for multiplayer poker", *Science*, 30 August 2019, https://science.sciencemag.org/content/early/2019/07/10/science.aay2400/tab-pdf.

and AI that is supposed to guide and guard safe practices when it comes to AI development and deployment.

More and more private companies all over the globe are publishing "ethical" frameworks, principles, guidelines, *etcetera*. Researchers at the Health Ethics & Policy Lab at ETH Zurich found that since 2016 there have been 84 documents containing ethical principles and guidelines released, most of which were produced by private actors.[17] These guidelines are, however, usually broad principles without concrete meaning given to terms like transparency, accountability, and non-maleficence. These ethical guidelines add to the awareness around the ethical implications of artificial intelligence, but don't successfully answer how companies play their part to remedy this. They also fail to shed light on how companies are held accountable should their behavior go against their guidelines. Although this phenomenon can be witnessed across the board in Big Tech[18], the most prominent example currently is Google's AI Principles.

Google has attracted lots of attention recently as employees have been critical of the corporation for a diverse range of its engagements: for example, on military contracts; disparities caused by their AI products; and the potential to assist authoritarian regimes. In light of controversial projects like Project Maven[19], Project Dragonfly[20], controversial board members, and other internal conflicts, Google's nascent AI ethics board only lasted one week before falling apart.[21] In the wake of employee protests and walk-outs, Google said

17 Anna Jobin, Marcello Ienca, and Effy Vayena, "Artificial Intelligence: the global landscape of ethics guidelines", Health Ethics & Policy Lab, ETH Zurich, 2019, https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf.

18 For example, Google, Apple, Facebook, Amazon and Microsoft (alongside many civil society actors) are all members of the Partnership on AI, a consortium seeking to shape best practices and public dialogue about AI's benefits for society, https://www.partnershiponai.org/partners/.

19 Devin Coldewey, "Google's new 'AI principles' forbid its use in weapons and human rights violations", *Tech Crunch,* 7 June 2018, https://techcrunch.com/2018/06/07/googles-new-ai-principles-forbid-its-use-in-weapons-and-human-rights-violations/?guccounter=2.

20 See the open letter to Google from Article19, a human rights watchdog focusing on the freedom of expression: Article 19, "Open letter: Response to Google on Project Dragonfly, China and human rights", 11 December 2018, https://www.article19.org/resources/open-letter-response-to-google-on-project-dragonfly-china-and-human-rights/.

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

it would conduct more transparent work in this realm going forward. However, a recent report by *The Intercept* indicates that instead of being party to government, military, or controversial contracts directly, Google has been providing financial, technological, and engineering support to many startups through a venture capital arm, Gradient Ventures. Instead of Google entering into contracts directly, the supported startups are the contracting parties entering into agreements with law enforcement or the military, for example.[22] While there may have been an effort from the tech giant to show transparent, responsible and ethical behavior by publishing ethical guidelines, it is not enough.

Public sector actors also cling to AI ethics as a way forward. EU institutions often talk about how a strategic approach to AI is needed, and the path some pursue involves being the leader in AI ethics. The European Commission put forward a European approach to AI in a Communication in the spring of 2018, which expresses the European goal of "leading the way in developing and using AI for good and for all, building on its values and strengths."[23] To do this, the EU set up the High-Level Expert Group on Artificial Intelligence (AI HLEG) to take the first steps.[24] Civil society actors have similarly set up watchdog organizations and reporting bodies[25] to raise awareness in potential ethical harms caused by AI applications. Research institutes have set up research consortiums to attract and gather the best minds to solve the problems at hand. The Beijing Academy of Artificial Intelligence (BAAI) (backed by the Chinese Ministry of Science) announced the Beijing AI Principles, which is a code of ethics for AI with guiding principles for research and development (R&D).[26] This declaration put China in the same group as France, Canada, and the UK – other countries who have developed national ethics guidelines.

22 Lee Fang, "Google continues investments in military and police AI technology through venture capital arm", *The Intercept*, 23 July 2019, https://theintercept.com/2019/07/23/google-ai-gradient-ventures/.

23 European Commission, "Communication Artificial Intelligence for Europe", Communication, 25 April 2018, https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.

24 See the HLEG's "Ethics guidelines for trustworthy AI" from the spring of 2019: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

25 Some organizations include AI Now Institute, AI Ethics Lab, and more: Mia Dand, "12 Organizations Saving Humanity from the Dark Side of AI", *Data Driven Investor*, 27 September 2018, https://medium.com/datadriveninvestor/12-organizations-saving-humanity-from-the-dark-side-of-ai-bce8c9da1ea5.

26 Will Knight, "Why does Beijing suddenly care about AI ethics?", *Technology Review*, 31 May 2019, https://www.technologyreview.com/s/613610/why-does-china-suddenly-care-about-ai-ethics-and-privacy/.

These diverse public sector initiatives show that ethics and AI governance today go hand in hand. But the focus on ethics limits effective governance that could ensure that the design and use of AI technologies do not continue to cause undue harm.

The ethics discourse thus far has certainly been an important first step to gain momentum, however it has had its limitations. The mushrooming of ethical initiatives and guidelines shows that actors appear to start from scratch with every iteration. There is no effective agreement between key players regarding how AI is to be designed ethically, upheld by developers and corporate actors, or enforced by government actors. On top of little consensus within respective domestic settings, there is also seldom international agreement on how to best regulate the development and use of new AI and potentially harmful AI technologies, as well as how to best enforce regulation on a technology that is as globalized and transborder in nature. Furthermore, there is seldom shared language between different stakeholders, regions, or sectors. This is why the discourse should move past "ethics" and anchor the debate in human rights. There is no international ethics regime, but there is an international human rights regime.

Chapter 3: Need to Tie the Ethical Debate to Human Rights

Raising awareness around the ethical implications of these technologies is a significant contribution from private, public, and third sector actors. These efforts are valuable and important. However, focusing on ethics has its limitations. One way to compliment the ethical debate is to anchor it in human rights. Ethical guidelines, principles, and values may be in vain because what exactly they *mean* and how to *enforce* them remains unclear. While they are a step in the right direction, they lack enforcement capabilities.[27] Ethics is a fig leaf that leads to impunity when it comes to international human rights. Leadership should be defined not by those who come up with the latest ethical frameworks, but by those who develop concrete obligations regarding AI that are rooted in and derived from a human rights perspecitve.

27 Article 19, "Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence", April 2019, https://www.article19.org/wp-content/uploads/2019/04/Governance-with-teeth_A19_April_2019.pdf.

The human rights framework provides a clear and widespread legal basis

The human rights framework has proven timeless and agile and can be applied to contemporary challenges. An example of the application of human rights to modern challenges can be seen with digital rights[28], or rights that allow us to access, use, create and publish digital media, access and use computers and communications networks, and dictate what data constitutes personal data and therefore the right to be kept private. Digital rights protect more than just privacy and also include freedom of assembly, association and expression.[29] After rapid digitalization and advances in information and communication technologies led to unforeseeable consequences for individuals' rights and freedoms, a new international framework to uphold digital rights was not created. Instead, digital rights lawyers and human rights lawyers simply took the existing conventions and treaties and said that human rights must be protected *online and offline*. This is, however, an ongoing battle. A lot of work still needs to be done to elucidate the scope of digital rights, and to ensure that rights and freedoms, "whether exercised in person, through technologies of today, or through technologies that will be invented in the future"[30], are protected by international law. These protections are also important to guide the development of artificial intelligence.

Developing AI with human rights standards and holding AI accountable to protect individuals' basic rights and freedoms addresses many of the limitations of ethics. With human rights, we already have international consensus on which we can build. Though not completely universal, the human rights framework is one of the most widespread international legal frameworks in existence, with United Nations Member States[31] bound by conventions and

28 United Nations Human Rights Council, "The promotion, protection and enjoyment of human rights on the Internet", Resolution 20/8, 16 July 2012.

29 United Nations Human Rights Council, "Right to freedom of peaceful assembly and of association", Report of the Special Rapporteur, A/HRC/41/41, 17 May 2019.

30 Ibid.

31 Depending on the specific treaties states have ratified.

treaties developed over the decades.[32] Human rights law is clear, has legal basis and precedents, and is internationally binding. These are three characteristics that would significantly benefit AI development, especially in recent years as it has caused increasingly more harm and irreversible damage.

Signatories not only have international obligations, but many have adopted these conventions into domestic constitutions[33] as well. Once ratified, states commit to upholding rights and freedoms for all by adopting human rights into their domestic law and by protecting, respecting and fulfilling the rights and freedoms of individuals within their jurisdictions – and this includes protection against harm that may be caused by third party actors. States need to be able to check the lawfulness of an accumulation power within private domestic enterprises.[34] This has however proven difficult. As Big Tech companies shape domestic activities around AI more and more each year, their political, societal, and economic might grows stronger. They should be held accountable by states[35], both domestically and internationally. Moreover, the technologies created by these private actors may allow states to govern in ways that infringe upon fundamental rights of citizens. States must similarly be held accountable by the international community. Because human rights are codified in international law and institutions, they can provide a framework to uphold accountability and responsibility in AI technologies in a way that is currently not being done effectively enough by the ethics discourse.

32 In the mid 20th Century, leaders from around the world came together to draft the Universal Declaration of Human Rights (UDHR). This non-binding text has inspired and led to many binding human rights conventions and accords, such as the International Covenant on Civil and Political Rights (ICCPR), the International Covenant on Economic, Social and Cultural Rights (ICESCR), and the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW) to name a few.There are now ten bodies within the UN who promote and protect human rights.

33 How the terms of these conventions are incorporated into domestic law is a function of the individual country's constitution.

34 UN Human Rights Council, "Report of the Working Group on the issue of human rights and transnational corporations and other business enterprises on the seventh Forum on Business and Human Rights", A/HRC/41/49, May 2019.

35 The French state, for example, attempted to hold Facebook accountable for its inability to crack down on dangerous content. Facebook failed to provide French regulators enough information about its algorithms and this "lack of transparency (...) justifies an intervention of public authorities." See the full report from Angela Charlton and the Associated Press, "France threatens new rules on Facebook as Zuckerberg visits", 10 May 2019, https://www.pbs.org/newshour/economy/france-threatens-new-rules-on-facebook-as-zuckerberg-visits.

Artificial intelligence can build on digital rights' legal precedents

The UN has frameworks and processes in place to protect the individual rights and freedoms of humans all over the world, and they can be applied to artificial intelligence similar to other digital issues. Emerging technologies such as machine learning applications tend to threaten not only civil and political rights (for example control or manipulation of content on social media sites violating the freedom of expression) protected by the International Covenant on Civil and Political Rights (ICCPR), but also social and economic rights (for example, increased use of AI risk-assesment tools in pre-trial release hearings violating the right to non-discrimination) protected by the International Covenant on Economic, Social and Cultural Rights (ICESCR).[36] As global connectivity increases, AI applications have the potential to become ubiquitous no matter the region or sector of the economy. Interaction with AI technologies in aspects of everyday life may increase for an individual, either knowingly or unbeknownst to her or him. This stands to make infringements of human rights even harder to combat.

Threats to human rights by artificial intelligence are abundant and diverse. Examples include algorithmic bias; AI for profiling, AI-powered advertising and targeting (violating the freedom from discrimination); the use of facial recognition software to supercharge a surveillance state; datafication of marginalized communities, growing digital footprints, fusion of data and availability of biometric data (violating the right to privacy[37]); AI for content takedowns; and systematic information control by state actors (violating the freedom of expression[38] and the freedom of peaceful assembly and association[39]); and many many more.[40] While these may be novel challenges to individual rights and freedoms, and while the scale and impact of these in-

36 Nani Jansen Reventlow, "Digital rights are *all* human rights, not just civil and political", *Medium*, Berkman Klein Center Collection, 27 February 2019, https://medium.com/berkman-klein-center/digital-rights-are-all-human-rights-not-just-civil-and-political-daf1f1713f7a.

37 Privacy International & Article 19, "Privacy and Freedom of Expression In the Age of Artificial Intelligence", April 2018, https://www.privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf.

38 Ibid.

39 UN Human Rights Council, "Rights to freedom of peaceful assembly and of associate", Report, A/HRC/41/41, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/141/02/PDF/G1914102.pdf?OpenElement.

40 United Nations General Assembly, "Promotion and protection of the right to freedom of opinion and expression", 29 August 2018, https://undocs.org/A/73/348.

fringements may be greater due to the pervasiveness and speed of ML technologies, they are still iterations of traditional human rights violations. Why not take the strategy applied by digital rights lawyers and fight for human rights online and offline in the case of AI technologies as well?

If AI technologies were subject to scrutiny on the international stage as violations of human rights, rather than just harmful and unethical, then there would be more enforcement mechanisms with which to regulate the key innovators and perpetrators. In addition to ethical, responsible, or trustworthy frameworks for design and use, AI needs a framework for development and use in accordance with human rights. Human rights specialists need to work in conjunction with technologists, and machine learning technologies need to be scrutinized more by human rights lawyers, states, and international institutions. Industry and private sector actors responsible for the development, deployment, and curation of ML technologies must also be held accountable by this international framework. Whereas company guidelines may consider internal affairs at one company in particular, the UN Human Rights Office of the High Commissioner (OHCHR) Guiding Principles on Business and Human Rights[41] is one human rights-related framework that informs the debate on how to hold the private sector accountable under sovereign states' commitments to international human rights. This UN framework for holding private actors accountable for human rights violations says that while states are not directly responsible for human rights violations caused by private actors within their territory, they are responsible if they do not do everything in their power to prevent, punish and redress abuse by the private sector. This is reiterated in a report on human rights and transnational corporations from the UN Human Rights Council: describing a session held on AI, the report addresses the implications of the three pillars of the Guiding Principles (protect, respect, remedy) regarding emerging technologies, and emphasized that "both the tech companies engineering new products and companies deploying them in their business needed to exercise due diligence and consider how their actions might impact on [sic] human rights."[42]

41 United Nations Human Rights Office of the High Commissioner, "Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework", 2011, https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf.

42 UN Human Rights Council, "Report of the Working Group on the issue of human rights and transnational corporations and other business enterprises on the seventh Forum on Business and Human Rights", A/HRC/41/49, May 2019.

Globalized technologies require international obligations

Promoting a commitment to human rights in the artificial intelligence discourse can also foster international cooperation. Relying on local regulation[43] is simply fire-fighting – for example, while banning facial recognition in individual cities may be an attempt to redefine legality of AI ubiquity, it is significantly narrower in scope than focusing on the procurement of facial recognition software by state actors globally and what human rights this violates. If the international institutions exist, why not take advantage of them? Countries should focus on holding their private sector actors accountable and applying AI only in ways that respect constitutional obligations domestically, as well as international obligations to universal human rights. Putting international human rights at the center of the debate can also encourage international cooperation and combining the efforts of like-minded states and regions. For Europe in particular, banding together with the region and using this collective force as a way to hold other states accountable is more promising than trusting single domestic legislations to go it alone. Refocusing on a global framework like human rights law can serve as a unifying force and bring together states of all regions to hold actors accountable for the development and deployment of AI technologies.

Anchoring the AI discourse in human rights opens the door to utilize UN institutions for this purpose. With bodies, multilateral organizations, regional treaties, and more, there are ample opportunities for states to convene and map a common way forward. States need to use international fora to not only raise awareness of the human rights implications of artificial intelligence globally, but also to synergize efforts to prevent abuses, protect citizens, and provide effective judicial or non-judicial remedy for individuals whose rights have been abused – both by states and private actors.

Engaging at the UN level, for example, in the UN Human Rights Council – the only forum for governmental dialogue around human rights – would be a first step. This body launches fact-finding missions, investigates certain cases, and has a rotating periodic review mechanism to focus investigations on different countries. These tools could be useful to raise awareness around

43 See for example the San Francisco city ban on facial recognition technologies (see: Dave Lee, "San Francisco is first US city to ban facial recognition", *BBC News*, 15 May 2019,https://www.bbc.com/news/technology-48276660), as well as other cities to follow in San Francisco's footsteps (see: Rachel Metz, "Beyond San Francisco, more cities are saying no to facial recognition", *CNN Business*, 17 July 2019,https://edition.cnn.com/2019/07/17/tech/cities-ban-facial-recognition/index.html).

human rights violations caused by AI technologies and call for more accountability for state and private actors. This body has, however, come with its fair share of criticism: the Trump Administration pulled the US out of the Council in 2018; China remains a member despite the loud calls of advocates to remove them, saying that the state is trying to push a national agenda based on the notion of sovereignty[44], going against the institution; and other states who have been accused of human rights abuses sit on the Council.[45] Though this is the most prominent Charter-based body of the UN focusing on human rights, there are other bodies within the UN that promote and protect human rights. The United Nations General Assembly, the Economic and Social Council, and the International Court of Justice, among others[46], could be leveraged for convening power, awareness raising, and international cooperation around AI going forward.

While momentum is building, efforts have been few and far between, rather than streamlined and comprehensive. States and companies alike need to reiterate and underline their commitments to international human rights law and embrace international conventions that are currently in place to avoid more widespread damage to basic rights and freedoms in the wake of AI development and use. Human rights must be accepted as a minimum requirement, as that is a clear and concrete signal for *how* human rights should be used and applied to contemporary problems.[47] The global tensions surrounding quick and competitive development of AI systems, as well as adoption of AI for authoritarian means are extremely harmful for individuals and their rights and freedoms. It is important that international cooperation rather than competition is fostered. Embracing international conventions and declarations committing states to uphold the international human rights legal framework is one way to foster this cooperation.

44 Lindsay Maizland, "Is China Undermining Human RIghts at the United Nations?", *Council on Foreign Relations*, 9 July 2019, https://www.cfr.org/in-brief/china-undermining-human-rights-united-nations.

45 Krishnadev Calamur, "The UN Human Rights Council Is a Deeply Flawed Body", *The Atlantic*, 20 June 2018, https://www.theatlantic.com/international/archive/2018/06/us-un-human-rights/563276/.

46 See a list of all UN bodies involved in human rights promotion and protection: https://www.ohchr.org/EN/HRBodies/Pages/OtherUnitedNationsBodies.aspx

47 Article 19, "Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence", April 2019, https://www.article19.org/wp-content/uploads/2019/04/Governance-with-teeth_A19_April_2019.pdf.

Conclusion

Centering the AI discourse around human rights rather than simply ethics can be one way of providing a clearer legal basis for development and deployment of AI technologies. However, this does not spare us from the hard work that it will require to flesh out what exactly the human rights implications of AI are. Because AI and ML applications are extremely diverse in context and impact, it also remains far from clear how AI may affect human rights. We need more analysis and research on specific case studies and contexts around how AI can and must respect human rights.

The road ahead will be challenging. The US and China, arguable the most dominant actors in the AI realm, are both criticized for diverse human rights abuses on their own soil[48]; holding these major innovators accountable for violations of human rights, especially those caused or exacerbated by AI technologies, will be demanding. There are also stagnant conversations in Geneva, prohibiting any effective regulation on lethal autonomous weapons systems (LAWS) and whether or how they can be researched and adopted. Despite these challenges, however, the outlook is positive. There is a growing amount of activity in international fora, especially those concerned with human rights. Many stakeholders have begun to engage in and strengthen the human rights-centered AI discourse. The United Nations has been instrumental in pushing for the application of the human rights legal framework to AI development and deployment. The General Assembly has published reports from the Special Rapporteur David Kaye on promoting specific rights and freedoms that may be threatened by AI[49]; reports from the UN High Commissioner for Human Rights to inform the Human Rights Council on identifying and clarifying principles, standards and best practices to promote the right to privacy in the digital age[50]; draft resolutions that uphold certain

48 Human Rights Watch, "World Report 2019: China", 2019, https://www.hrw.org/world-report/2019/country-chapters/china-and-tibet.
Human Rights Watch, "World Report 2019: United States", 2019, https://www.hrw.org/world-report/2019/country-chapters/united-states.

49 See for instance the UN General Assembly, "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression", A/73/348, 29 August 2018, https://undocs.org/A/73/348.

50 See the UN General Assembly, "The right to privacy in the digital age: Report of the United Nations High Commissioner for Human Rights", A/HRC/39/29, 3 August 2018, https://undocs.org/A/HRC/39/29.

rights affected by AI and the digital age[51]; and reports from the Working Group on the issue of human rights and transnational corporations, focusing on technology and corporate respect for human rights.[52] As the convening body of human rights conventions, treaties and declarations, the appropriate UN mechanisms and procedures should continue to push this legal framework and scrutinize Member States and corporate actors that fail to uphold their international and constitutional obligations and allow the human rights of citizens to be infringed upon by AI technologies.

Another promising outlook comes from civil society. Article 19, a human rights organization that promotes specifically the right to freedom of expression and information, has also released reports on the negative effects of AI on this freedom. In one report from April 2019, "Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence", Article 19 breaks down the societal impact of AI into two dimensions: the normative approach and the technical approach. While the normative approach addresses the ethical initiatives and gaps, the technical approach takes a much needed step. The report lays out how communities of researchers, academics and scientists have been working to make AI systems that are fair, accountable, and transparent (FAT). This report shows that the human rights-centered AI discourse must take place not only among the legal community, ethicists, or social scientists but can also inform the scientific community, technologists, and engineers and be part of future standards of AI development.

This paper does not seek to say that a human rights-centered AI discourse is without limitations. The international community would still need to raise awareness, build consensus, and flesh out exactly how AI technologies violate human rights, and develop ways for effective legal remedies. Even considering these limitations, however, focusing the discourse on human rights rather than just ethical guidelines provides more accountability measures, more obligations for state and private actors, and relies on clear, consistent and widely accepted legal principles that developed over time. Reaching a common understanding of how AI needs to respect human rights is a much more coherent route to further encourage the adoption of broadly applicable principles, as compared to continuing with the current proliferation of vague

51 See also the UN General Assembly, "The right to privacy in the digital age", revised draft resolution, A/C.3/73/L.49/Rev.1, 14 November 2018, https://undocs.org/A/C.3/73/L.49/Rev.1.

52 UN Human Rights Council, "Report of the Working Group on the issue of human rights and transnational corporations and other business enterprises on the seventh Forum on Business and Human Rights", A/HRC/41/49, May 2019.

ethical guidelines. Although not perfect, it is a step that the international community must take in order to ensure that the development and deployment of AI technologies does not continue to infringe on fundamental rights of citizens across the globe.

About the Stiftung Neue Verantwortung

The Stiftung Neue Verantwortung (SNV) is an independent think tank that develops concrete ideas as to how German politics can shape technological change in society, the economy and the state. In order to guarantee the independence of its work, the organisation adopted a concept of mixed funding sources that include foundations, public funds and businesses.

Issues of digital infrastructure, the changing pattern of employment, IT security or internet surveillance now affect key areas of economic and social policy, domestic security or the protection of the fundamental rights of individuals. The experts of the SNV formulate analyses, develop policy proposals and organise conferences that address these issues and further subject areas.

About the Authors

Kate Saslow works as project manager for Artificial Intelligence and Foreign Policy; EU Cyber Resilience; and Machine Learning and Cyber Security. Kate holds a Master of International Affairs degree from the Hertie School of Governance. She specialized in topics of digitalization, economic development, and governance, and concluded her studies with a capstone thesis on artificial intelligence and gender inequality in the United States labor economy. She also holds a Bachelor's degree in International Relations from the University of Michigan, Ann Arbor.

Philippe Lorenz leads the SNV's Artificial Intelligence and Foreign Policy project, assessing the implications of artificial intelligence for international relations. In his work, Philippe Lorenz focuses on policy areas that are undergoing major changes due to advanced technologies. Philippe studied law at the University of Passau and International Relations at the Rhine-Waal University of Applied Sciences in Kleve.

How to Contact the Authors

Kate Saslow
ksaslow@stiftung-nv.de
+49 (0)30 81 45 03 78 84

Philippe Lorenz
plorenz@stiftung-nv.de
+49 (0)30 81 45 03 78 94

Imprint

Stiftung Neue Verantwortung e. V.

Beisheim Center
Berliner Freiheit 2
10785 Berlin

T: +49 (0) 30 81 45 03 78 80
F: +49 (0) 30 81 45 03 78 97

www.stiftung-nv.de
info@stiftung-nv.de

Design:
Make Studio
www.make-studio.net

Layout:
Johanna Famulok

Free Download:

www.stiftung-nv.de