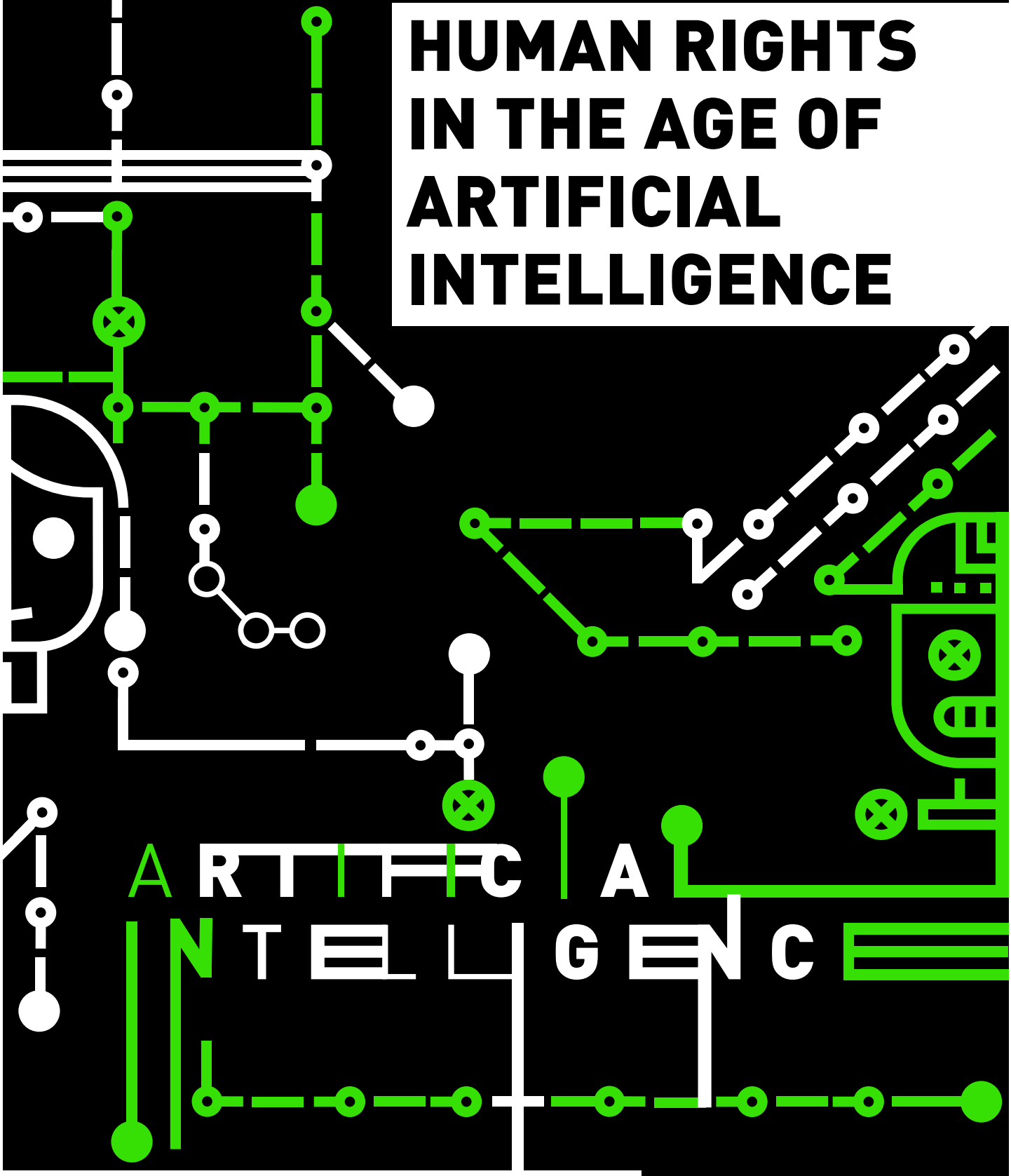
accessnow.org

HUMAN RIGHTS IN THE AGE OF ARTIFICIAL INTELLIGENCE



Access Now defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age.

accessnow

This report is a product of Access Now. We thank lead author Lindsey Andersen for her significant contributions. If you have questions about this report or you would like more information, you can contact **info@accessnow.org.**

CONTENTS

HUMAN RIGHTS IN THE AGE OF ARTIFICIAL INTELLIGENCE

I. EXECUTIVE SUMMARY

As artificial intelligence continues to find its way into our daily lives, its propensity to interfere with human rights only gets more severe. With this in mind, and noting that the technology is still in its infant stages, Access Now conducts this preliminary study to scope the potential range of human rights issues that may be raised today or in the near future.

Many of the issues that arise in examinations of this area are not new, but they are greatly exacerbated by the scale, proliferation, and real-life impact that artificial intelligence facilitates. Because of this, the potential of artificial intelligence to both help and harm people is much greater than from technologies that came before. While we have already seen some of these consequences, the impacts will only continue to grow in severity and scope. However, by starting now to examine what safeguards and structures are necessary to address problems and abuses, the worst harms—including those that disproportionately impact marginalized people—may be prevented and mitigated.

There are several lenses through which experts examine artificial intelligence. The use of international human rights law and its well-developed standards and institutions to examine artificial intelligence systems can contribute to the conversations already happening, and provide a universal vocabulary and forums established to address power differentials.

Additionally, human rights laws contribute a framework for solutions, which we provide here in the form of recommendations. Our recommendations fall within four general categories: data protection rules to protect rights in the data sets used to develop and feed artificial intelligence systems; special safeguards for government uses of artificial intelligence; safeguards for private sector uses of artificial intelligence systems; and investment in more research to continue to examine the future of artificial intelligence and its potential interferences with human rights.

Our hope is that this report provides a jumping off point for further conversations and research in this developing space. We don't yet know what artificial intelligence will mean for the future of society, but we can act now to build the tools that we need to protect people from its most dangerous applications. We look forward to continuing to explore the issues raised by this report, including through work with our partners as well as key corporate and government institutions.

II. INTRODUCTION

The concept of artificial intelligence has been elevated from the realm of science fiction to discussions in the highest circles of academia, industry, and government. However, experts have only just begun to look at the impact of artificial intelligence on human rights, and so far they do not even seem to agree on what the term means.

It is evident that use of artificial intelligence and machine learning technology has the potential to effect revolutionary changes in the world. In 2018, it was a key topic at RightsCon, Access Now's annual conference on the intersection of human rights and technology. Leading up to RightsCon, we worked with close partners to draft and publish the *Toronto Declaration on protecting the rights to equality and non-discrimination in machine learning systems*.[1] We also participated in a workshop on artificial intelligence and human rights hosted by the Data & Society Research Institute in New York, the goal of which was "to consider the value of human rights in the AI space, foster engagement and collaboration across sectors, and develop ideas and outcomes to benefit stakeholders working on this issue moving forward."[2]

This report is a preliminary scoping of the intersection of artificial intelligence and human rights. The first section proposes definitions for key terms and concepts, including "artificial intelligence" and "machine learning." We next look at how different artificial intelligence systems are used in the world today and ways in which they can both help or harm society. Turning to human rights, we look at the role human rights law can play in the development of artificial intelligence, including the interplay between these fundamental rights and ethics. Then, looking at widely adopted human rights instruments, we highlight the ways current and foreseeable uses of artificial intelligence can interfere with a broad range of human rights. Finally, we offer a list of recommendations for stakeholders to protect those rights.

We recognize that we are offering recommendations in the early stages of the development and use of artificial intelligence, and we are only beginning to grapple with its potential consequences. That is why one of our recommendations is to direct additional funding and resources to investigate further the issues raised in this report to determine what the safeguards and structures should be for preventing or mitigating future human rights abuses.

1 https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf.
2 https://points.datasociety.net/artificial-intelligence-human-rights-a-workshop-at-data-society-fd6358d72149.

III. DEFINITIONS

1. ARTIFICIAL INTELLIGENCE OR AI: There is no agreed-upon definition of artificial intelligence. Marvin Minsky, one of the founding AI scholars, defines it as "the science of making machines do things that would require intelligence if done by men."[3] Another founding scholar, John McCarthy, defines it as "the science and engineering of making intelligent machines."[4] A recent Stanford University report defines AI as "a science and a set of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems and bodies to sense, learn, reason, and take action."[5]

Stuart Russell and Peter Norving, authors of a popular AI textbook, suggest that AI can be broken down into the following categories: 1) systems that think like humans; 2) systems that act like humans; 3) systems that think rationally; and 4) systems that act rationally.[6]

In reality, AI is considered more of a field than an easily definable "thing," and it can be broken down into many subfields, such as machine learning, robotics, neural networks, vision, natural language processing, and speech processing. There is significant crossover among these sub-fields. AI also draws from fields other than computer science, including psychology, neuroscience, cognitive science, philosophy, linguistics, probability, and logic.

"Narrow AI"—what is currently in use—is the single-task application of artificial intelligence for uses such as image recognition, language translation, and autonomous vehicles. Machines currently perform more accurately than humans at these types of tasks. In the future, researchers hope to achieve "artificial general intelligence" (AGI). This would involve systems that exhibit intelligent behavior across a range of cognitive tasks. However, researchers do not estimate that these capabilities will be achieved for at least decades.[7]

2. BIG DATA: Datasets that are too large or complex for traditional data processing software to analyze. The increasing availability of big data, thanks to society's ever-expanding internet use, and coupled with rapid improvements in computing power, has enabled the significant advances in AI in the past 10 years.

3. DATA MINING: The process of discovering patterns and extracting information from large datasets. In the era of big data, data mining is often facilitated by machine learning.

4. MACHINE LEARNING (ML): Machine learning is a sub-field of AI. Harry Surden defines machine learning as "computer algorithms that have the ability to "learn" or improve in performance over time on some task."[8] Essentially, it is a machine that learns from data over time. This learning is through "a statistical process that starts with a body of data and tries to derive a rule or procedure that explains the data or can predict future data."[9] The resulting output is called a model. This is different from the traditional approach to artificial intelligence, which involved a programmer trying to translate the way humans make decisions into software code. The vast majority of artificial intelligence in the world today is powered by machine learning. Currently, many ML systems are far more accurate than humans at a variety of tasks, from driving to diagnosing certain diseases.[10]

3 "Report of COMEST on Robotics Ethics; 2017," n.d., 17.
4 McCarthy, John. 2018. " What Is AI? / Basic Questions "Jmc.Stanford.Edu. Accessed June 15 2018. http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.
5 2018. Ai100.Stanford.Edu. Accessed June 15 2018. https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf.
6 Qtd. in Committee on Technology, National Science and Technology Council, "Preparing for the Future of Artificial Intelligence" (Executive Office of the President of the United States, October 2016), 5, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.
7 "Preparing for the Future of Artificial Intelligence," 7.
8 Surden, Harry. 2014. "Machine Learning And Law". Papers.Ssrn.Com. Accessed June 15 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2417415.
9 "Preparing for the Future of Artificial Intelligence", 5.
10 For a visual explanation of how machine learning works, see, http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Machine learning works like this:[11]

(1) **Programmers begin with a historical data set**, which is divided into a training set and a test set.
(2) **They then choose a model**, a mathematical structure that characterizes a range of possible decision-making rules. This model includes adjustable parameters. The model is like a box, and the parameters are the adjustable knobs on the box.
(3) **They define an objective function** used to evaluate the desirability of the outcome.
(4) **They train the model**, which is the process of adjusting the parameters to maximize the objective function.
(5) Once trained, **they use the test dataset to evaluate the accuracy and effectiveness of the model.** Ideally, the model should perform similarly on the test data. The goal is to be able to generalize the model, so it is accurate in response to cases it has never seen before.

5. DEEP LEARNING: A machine learning technique that uses structures called "neural networks" that are inspired by the human brain. These consist of a set of layered units, modeled after neurons. Each layer of units processes a set of input values and produces output values that are passed onto the next layer of units. Neural networks often consist of more than 100 layers, with a large number of units in each layer to enable the recognition of extremely complex and precise patterns in data.

To explore this further, consider image recognition software that utilizes neural networks to identify a picture of an elephant. The first layer of units might look at the raw image data for the most basic patterns, perhaps that there is a thing with what appears to be four legs. Then the next layer would look for patterns within these patterns, perhaps that it is an animal. Then maybe the next layer identifies the trunk. This process would continue throughout many layers, recognizing increasingly precise patterns in the image, until the network is able to identify that it is indeed a picture of an elephant.

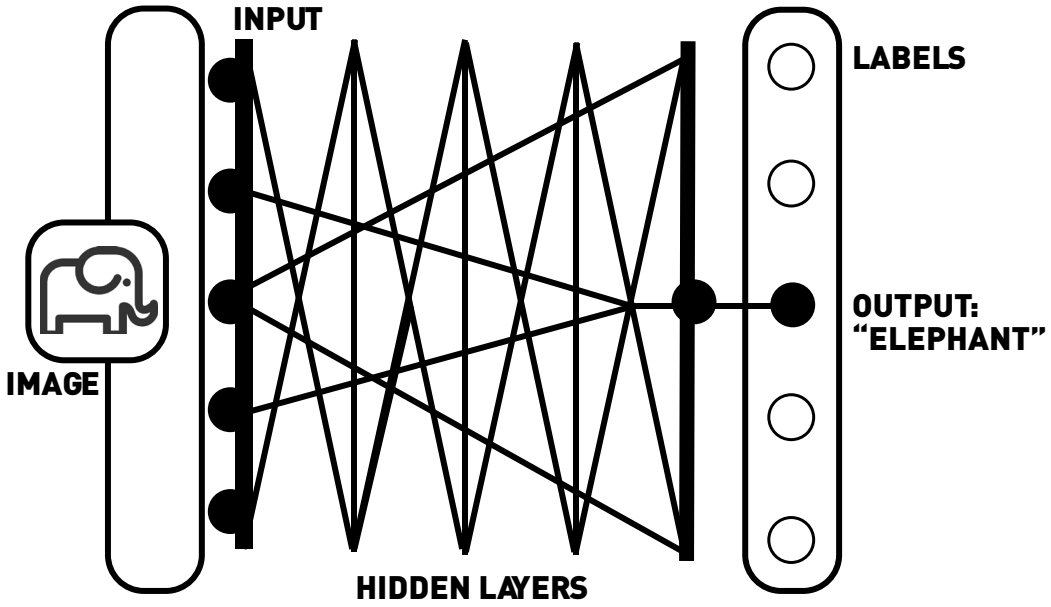


FIG 1. The 'neural networks' structure

11 "Preparing for the Future of Artificial Intelligence", 9.

Progress in deep learning has been the cause for much of the optimism about AI due its ability to process and find patterns in massive amounts of data with accuracy.[12] Whereas early ML typically uses a decision-tree structure, deep learning has become the dominant technique. It is often used to power specific ML approaches such as machine vision and natural language processing.[13]

5.1. MACHINE VISION: A specific ML approach that allows computers to recognize and evaluate images.[14] It is used by Google to help you search images and by Facebook to automatically tag people in photos.

5.2. NATURAL LANGUAGE PROCESSING: A specific ML approach that helps computers understand, interpret, and manipulate human language. It does this by breaking down language into shorter pieces and discovering how the pieces fit together to create meaning. Natural language processing enables commonly used services such as Google Translate and chatbots.[15]

5.3 SPEECH RECOGNITION: A specific ML approach allows computers to translate spoken language into text.[16] It allows you to use talk-to-text on your smartphone. It is often paired together with natural language processing and is used to power virtual assistants like Siri and Alexa.

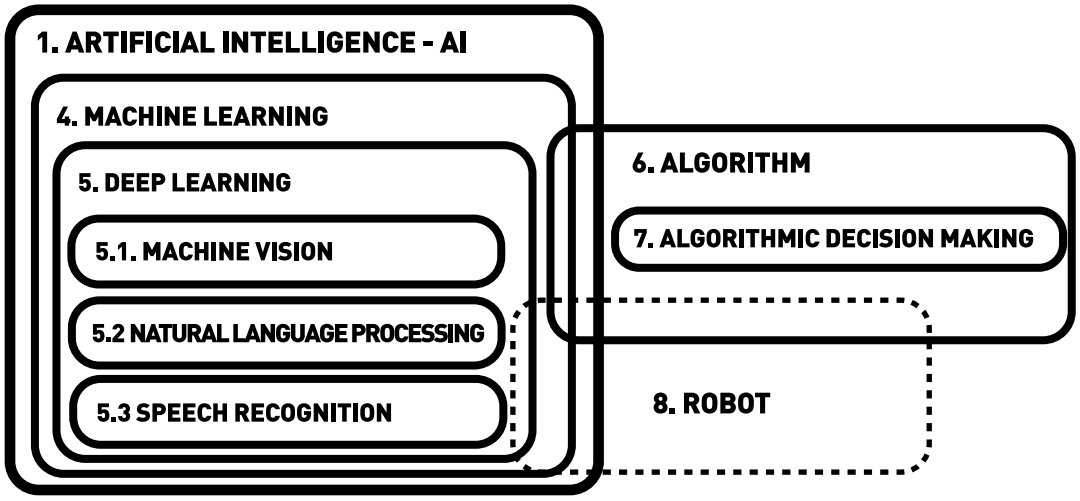


FIG 2. The relationships of the concepts in artificial intelligence

6. ALGORITHM: At its most simple, an algorithm is "a set of guidelines that describe how to perform a task."[17] Within computer science, an algorithm is a sequence of instructions that tell a computer what to do.[18] AI works through algorithms (neural networks are a type of algorithm), but not all algorithms involve artificial intelligence.

12 Preparing for the Future of Artificial Intelligence", 9–10.
13 Virtual assistants Siri, Cortana, Alexa use neural networks for speech recognition and to imitate human conversation. Deep learning in this case allows these virtual assistants to detect and understand the nuances of speech and produce a response that feels conversational. See https://machinelearning.apple.com/2017/08/06/siri-voices.html for more information. IBM's Watson uses deep learning techniques for machine vision to analyze to quickly and accurately interpret medical information and help doctors diagnose disease. See https://www.ibm.com/watson/health/ for more information.
14 "What Is a Machine Vision System (MVS)? - Definition from Techopedia," Techopedia.com, accessed May 12, 2018, https://www.techopedia.com/definition/30414/machine-vision-system-mvs.
15 " "What Is Natural Language Processing?," accessed May 12, 2018, https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html.
16 "Speech Recognition," Wikipedia, May 1, 2018, https://en.wikipedia.org/w/index.php?title=Speech_recognition&oldid=839191878.
17 "What Is an Algorithm? An Explainer.," accessed May 12, 2018, http://www.slate.com/articles/technology/future_tense/2016/02/what_is_an_algorithm_an_explainer.html.
18 Ibid.

7. ALGORITHMIC DECISION-MAKING: Using outputs produced by algorithms to make decisions. One of the earliest forms of algorithmic decision-making that is still in use today in the United States is federal sentencing guidelines for judges. This involves nothing more than a weighted mathematical equation, drawn from statistics, that recommends a sentence length based on the attributes of the crime.[19]

8. ROBOT: Robots often use many of the forms of artificial intelligence described above. However, by definition, robots have a physical body and mobility. Robots that use AI are able to perceive changes in their environment and function accordingly.[20] Although robots typically come to mind when thinking about artificial intelligence, they currently constitute a very small amount of our interactions with AI. AI in the field of robotics is a growing area of research and development but has not yet made nearly as many advancements or become as ubiquitous as non-robotic forms of machine learning.[21]

9. BOTS: Software applications that run automated tasks. Bots are increasingly being powered by ML, particularly chatbots, which use natural language processing to conduct human-like conversations with users.

10. OPEN DATA: Data that is freely available for everyone to view, use, share, and re-publish without restrictions. There is a broad open data movement that advocates that data should generally be treated this way.[22] In the context of AI, many advocates suggest training data for ML systems be open in order to surface bias and errors, as well as to shed light on the outputs ML systems produce. There is controversy on the best method to do this while respecting the privacy interests of data subjects.

11. PROTECTED INFORMATION: Information that includes, reflects, arises from, or is about a person's communications, and that is not readily available and easily accessible to the general public. While it has long been agreed that communications content deserves significant protection in law because of its capability to reveal sensitive information, it is now clear that other information arising from communications—metadata and other forms of non-content data—may reveal even more about an individual than the content itself, and thus deserves equivalent protection. Today, each of these types of information might, taken alone or analyzed collectively, reveal a person's identity, behavior, associations, physical or medical conditions, race, color, sexual orientation, national origins, or viewpoints; or enable the mapping of the person's location, movements or interactions over time, or of all people in a given location, including around a public demonstration or other political event. [23]

13. BIAS: There are both societal and statistical definitions of bias that come into play in AI. The societal definition of bias is "an inclination or prejudice for or against a person or group, especially in a way that is considered to be unfair."[24] The statistical definition of bias is the difference between the estimated—or predicted—value and the true value. In other words, the difference between what a system predicts and what actually happens.[25] In many cases, statistical bias present in a given AI system results in outcomes that are societally biased.

19 See https://www.ussc.gov/guidelines/2016-guidelines-manual/2016-chapter-5 for more information
20 "Report of COMEST on Robotics Ethics; 2017."
21 Raghav Bharadwaj, "Artificial Intelligence in Home Robots – Current and Future Use-Cases," TechEmergence, February 5, 2018, https://www.techemergence.com/artificial-intelligence-home-robots-current-future-use-cases/.
22 See https://en.wikipedia.org/wiki/Open_data and https://theodi.org/article/what-is-open-data-and-why-should-we-care/ for more info about the Open Data movement.
23 International Principles on the Application of Human Rights to Communications Surveillance, accessed June 15 2018, available at. https://necessaryandproportionate.org/.
24 https://en.oxforddictionaries.com/definition/bias
25 For a deeper discussion on statistical bias and fairness issues in AI, see talk by Princeton Computer Scientist Arving Narayanan: https://www.youtube.com/watch?v=jlXluYdnyyk

HOW DOES BIAS PLAY OUT IN AI?

AI can be biased both at the system and the data or input level. Bias at the system level involves developers building their own personal biases into the parameters they consider or the labels they define. Although this rarely occurs intentionally, unintentional bias at the system level is common. This often occurs in two ways:

- When developers allow systems to conflate correlation with causation. Take credit scores as an example. People with a low income tend to have lower credit scores, for a variety of reasons. If an ML system used to build credit scores includes the credit scores of your Facebook friends as a parameter, it will result in lower scores among those with low-income backgrounds, even if they have otherwise strong financial indicators, simply because of the credit scores of their friends.
- When developers choose to include parameters that are proxies for known bias. For example, although developers of an algorithm may intentionally seek to avoid racial bias by not including race as a parameter, the algorithm will still have racially biased results if it includes common proxies for race, like income, education, or postal code.[26]

Bias at the data or input level occurs in a number of ways:[27]

- The use of historical data that is biased. Because ML systems use an existing body of data to identify patterns, any bias in that data is naturally reproduced. For example, a system used to recommend admissions at a top university that uses the data of previously admitted students to train the model is likely to recommend upper class males over women and traditionally underrepresented groups.
- When the input data are not representative of the target population. This is called selection bias, and results in recommendations that favor certain groups over another. For example, if a GPS-mapping app used only input data from smartphone users to estimate travel times and distances, it could be more accurate in wealthier areas of cities that have a higher concentration of smartphone users, and less accurate in poorer areas or informal settlements, where smartphone penetration is lower and there is sometimes no official mapping.
- When the input data are poorly selected. In the GPS mapping app example, this could involve including only information related to cars, but not public transportation schedules or bike paths, resulting in a system that favored cars and was useless for buses or biking.
- When the data are incomplete, incorrect, or outdated. If there is insufficient data to make certain conclusions, or the data are out of date, results will naturally be inaccurate. And if a machine learning model is not continually updated with new data that reflects current reality, it will naturally become less accurate over time.

Unfortunately, biased data and biased parameters are the rule rather than the exception. Because data are produced by humans, the information carries all the natural human bias within it. Researchers have begun trying to figure out how to best deal with and mitigate bias, including whether it is possible to teach ML systems to learn without bias;[28] however, this research is still in its nascent stages. For the time being, there is no cure for bias in AI systems.

26 "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy: Cathy O'Neil: 9780553418811: Amazon.Com: Books," 155–60, accessed May 13, 2018, https://www.amazon.com/Weapons-Math-Destruction-Increases-Inequality/dp/0553418815.
27 Executive Office of the President of the United States, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," May 2016, 7–8, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.
28 This is broadly known at the FATML community, "Fairness, Accountability and Transparency for Machine Learning." See https://www.fatml.org/ for more info.

IV. WHAT MAKES THE RISKS OF AI DIFFERENT?

Many of the problems and risks explored in this report are not new. So how is AI different from the technologies that have come before it? Due to the ways AI has evolved from existing technologies, including in terms of both sophistication and scale, AI may exacerbate existing questions and introduce new problems to consider, with huge impacts for accountability and reliability. To illustrate this, consider two recent tech trends: big data and the rise of algorithmic decision-making.

Today, algorithmic decision-making is largely digital. In many cases it employs statistical methods similar to those used to create the pen-and-paper sentencing algorithm that we discussed above. Before AI, algorithms were deterministic—that is, pre-programmed and unchanging. Because they are based in statistical modeling, these algorithms suffer from the same problems as traditional statistics, such as poorly sampled data, biased data, and measurement errors. But because they are pre-programmed, the recommendations they make can be traced.

The use of AI in algorithmic decision-making has introduced a new set of challenges. Because machine learning algorithms use statistics, they also have the same problems with biased data and measurement error as their deterministic predecessors. However, ML systems differ in a few key ways. First, whereas traditional statistical modeling is about creating a simple model in the form of an equation, machine learning is much more fine-tuned. It captures a multitude of patterns that cannot be expressed in a single equation. Second, unlike deterministic algorithms, machine learning algorithms calibrate themselves. Because they identify so many patterns, they are too complex for humans to understand, and thus it is not possible to trace the decisions or recommendations they make.[29] In addition, many machine learning algorithms constantly re-calibrate themselves through feedback. An example of this are e-mail spam filters, which continually learn and improve their spam detection capabilities as users mark email as spam.

Another issue is the impact of error rates. Because of their statistical basis, all ML systems have error rates. Even though in many cases ML systems are far more accurate than human beings, there is danger in assuming that simply because a system's predictions are more accurate than a human's, the outcome is necessarily better. Even if the error rate is close to zero, in a tool with millions of users, thousands could be affected by error rates. Consider the example of Google Photos. In 2015 Google Photos' image recognition software was found to have a terribly prejudicial and offensive error: it was occasionally labeling photos of black people as gorillas. Because the system used a complex ML model, engineers were unable to figure out why this was happening. The only "solution" they could work out to this "racist" ML was merely a band-aid: they removed any monkey-related words from the list of image tags. [30]

Now, imagine a similar software system used by U.S Customs and Border Patrol that photographs every person who enters and exits the U.S. and cross-references it with a database of photos of known or suspected criminals and terrorists. In 2016, an estimated 75.9 million people arrived in the United States.[31] Even if the facial recognition system was 99.9% accurate, the 0.1% error rate would result in 75,900 people being misidentified. How many of these people would be falsely identified as wanted criminals and detained? And what would the impact be on their lives? Conversely, how many known criminals would get away? Even relatively narrow error rates in cases such as these can have severe consequences.

29 "What Is The Difference Between Machine Learning & Statistical Modeling," accessed May 12, 2018, https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/.
30 "When It Comes to Gorillas, Google Photos Remains Blind | WIRED," accessed May 13, 2018, https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/.
31 https://www.ustravel.org/answersheet

The bottom line: The scale, proliferation, and real-life impact of AI demands attention.

The proliferation of AI in data analytics has come with the rise of big data. In her 2015 book *Weapons of Math Destruction*, data scientist Cathy O'Neil documented how algorithmic decision-making is now ubiquitous in the West, from assigning credit scores, to identifying the best candidates for a job position, to ranking students for college admissions.Today, these algorithmic decision-making systems are increasingly employing machine learning, and they are spreading rapidly. They have many of the same problems as traditional statistical analysis. However, the scale and reach of AI systems, the trend of rapid, careless deployment, the immediate impact they have on many people's lives, and the danger of societies viewing their outputs as impartial, pose a series of new problems.

V. HELPFUL AND HARMFUL AI

Every major technological innovation brings potential to advance or damage society. The data processing and analysis capabilities of AI can help alleviate some of the world's most pressing problems, from enabling advancements in diagnosis and treatment of disease, to revolutionizing transportation and urban living, to mitigating the effects of climate change. Yet these same capabilities can also enable surveillance on a scale never seen before, can identify and discriminate against the most vulnerable, and may revolutionize the economy so quickly no job retraining program can possibly keep up. And despite major strides in the development of AI, the so-called "artificial intelligence revolution" is only a decade old, meaning there are many unknown possibilities in what is to come.

Below we identify some of the ways AI is being used to help or harm societies. It is important to note that even the "helpful" uses of AI have potentially negative implications. For example, many applications of AI in healthcare pose serious threats to privacy and risk discriminating against underserved communities and concentrating data ownership within large companies. At the same time, the use of AI to mitigate harm may not solve underlying problems and should not be treated as a cure for societal ailments. For example, while AI may alleviate the need for medical professionals in underserved areas, it isn't providing the resources or incentives those professionals would need to relocate. Similarly, some of the use cases categorized as "harmful" came about as a result of good intentions, yet are causing significant harm.

HELPFUL AI ☺

Improving access to healthcare and predicting disease outbreaks: Already there have been significant advancements through the use of AI in disease diagnosis and prevention. AI is also being used to improve access to healthcare in regions where there is a lack of access.[32] Victims of disease outbreaks also benefit from the use of AI to enable health officials to intervene early to contain an outbreak before it starts.[33]

Making life easier for the visually impaired: Tools for image recognition are helping people who are visually impaired better navigate both the internet and the real world.[34]

Optimizing agriculture and helping farmers adapt to change: AI is combining information from global

32 IBM's Watson is being used in hospitals around the world to help doctors diagnose and treat disease. See https://www.ibm.com/watson/health/ for more information. Another example is Aajoh, a Nigerian start-up developing an AI system for remote medical diagnosis. Users share their symptoms via text, audio, and photographs, and Aajoh uses AI to provide possible diagnoses. See Stephen Timm, "6 Artificial Intelligence Startups in Africa to Look out For," Venture Burn, April 24, 2017, http://ventureburn.com/2017/04/five-artificial-intelligence-startups-africa-look-2017/?platform=hootsuite.

33 https://www.cnn.com/2018/03/06/health/rainier-mallol-tomorrows-hero/index.html

34 For some examples see Facebook's efforst to help the blind "see" Facebook: "Using Artificial Intelligence to Help Blind People 'See' Facebook," Facebook Newsroom, April 4, 2016, https://newsroom.fb.com/news/2016/04/using-artificial-intelligence-to-help-blind-people-see-facebook/. See also Microsoft's work: "Seeing AI: An app for visually impaired people that narrates the world around you," Microsoft, https://www.microsoft.com/en-us/garage/wall-of-fame/seeing-ai/.

satellite imagery with weather and agronomic data to help farmers improve crop yields, diagnose and treat crop disease, and adapt to changing environments. This approach to farming is known as precision agriculture, and it can help increase farm productivity to feed more of the world's growing population.

Mitigating climate change, predicting natural disasters, and conserving wildlife: With the effects of climate change appearing around the world, machine learning is being used to make more accurate climate models for scientists. Already, AI is used to rank climate models and predict extreme weather events,[35] as well as to better predict extreme weather events and respond to natural disasters.[36] AI is also helpful for identifying and apprehending poachers and locating and capturing disease-spreading animals.

Making government services more efficient and accessible: Despite often being slow to adopt new technologies, governments around the world are using AI, from the local to the national levels, to make public services more efficient and accessible, with an emphasis on developing "smart cities." AI is also being used to allocate government resources and optimize budgets.[37]

HARMFUL AI ☹

Perpetuating bias in criminal justice: There are many documented cases of AI gone wrong in the criminal justice system. The use of AI in this context often occurs in two different areas: risk scoring—evaluating whether or not a defendant is likely to reoffend in order to recommend sentencing and set bail—or so-called "predictive policing," using insights from various data points to predict where or when crime will occur and direct law enforcement action accordingly.[38] In many cases, these efforts are likely well-intentioned. Use of machine learning for risk scoring of defendants is advertised as removing the known human bias of judges in their sentencing and bail decisions.[39] And predictive policing efforts seek to best allocate often-limited police resources to prevent crime, though there is always a high risk of mission creep.[40] However, the recommendations of these AI systems often further exacerbate the very bias they are trying to mitigate, either directly or by incorporating factors that are proxies for bias.

Facilitating mass surveillance: Given that AI provides the capacity to process and analyze multiple data streams in real time, it is no surprise that it is already being used to enable mass surveillance around the world.[41] The most pervasive and dangerous example of this is use of AI in facial recognition software.[42] Although the technology is still imperfect, governments are looking to facial recognition technology as a

35 https://www.scientificamerican.com/article/how-machine-learning-could-help-to-improve-climate-forecasts/
36 Aili McConnon, "AI Helps Cities Predict Natural Disasters," The Wall Street Journal, June 26, 2018, https://www.wsj.com/articles/ai-helps-cities-predict-natural-disasters-1530065100.
37 See Hila Mehr, "Artificial Intelligence for Citizen Services and Government," Ash Center for Democratic Governance and Innovation, Harvarf Kennedy School, August 2017, and https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf, and IBM Cognitive Business, "Watson helps cities help citizens: the 411 on how artificial intelligence transforms 311 services," Medium, January 31, 2017, https://medium.com/cognitivebusiness/watson-assists-cities-with-311-3d7d6898d132.
38 A 2016 ProPublica investigation revealed that not only was COMPAS, an ML-powered software widely used in the U.S. criminal justice system, was inaccurate at forecasting future crime and heavily biased against black defendants. The investigators looked at risk scores of over 7,000 people arrested in Broward County, Florida and compared them with subsequent criminal records. They found that only 20% of the people predicted to commit violent crimes went on to do so. And when looking at the full range of crimes, only 61% of defendants deemed likely to reoffend were actually arrested for a future crime. Jeff Larson Julia Angwin, "Machine Bias," text/html, ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
39 An investigation by the Science and Technology Committee of Parliament of HART, ML-powered software being used by police in Durham, England to evaluate recidivism risk, revealed that it was calibrated to avoid false negatives, incorrectly classifying a person as low risk when they in fact go on to commit serious crimes. https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/ and https://www.bbc.co.uk/news/technology-43428266
40 Public records suggest that software developed by Palantir and used by police in criminal investigations in New Orleans was used beyond its original intended scope. After a series of investigative reports and significant public outcry, the city ended its six-year contract with Palantir in March 2018. https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd and https://www.nola.com/crime/index.ssf/2018/03/palantir_new_orleans_nopd.html
41 China, in particular, is aggressively pursuing an AI-based surveillance state. See Paul Mozur, "Inside China's Dystopian Dreams: AI, Shame and Lots of Cameras," The New York Times, July 8, 2018, https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html.
42 In 2018, Australia unveiled a plan to connect its network of CCTV cameras to existing facial recognition and biometric databases. The proposed measure is pending in Parliament. Ihttps://www.accessnow.org/cms/assets/uploads/2018/07/Human-Rights-in-the-Digital-Era-an-international-perspective-on-Australia.pdf.

tool to monitor their citizens, facilitate profiling of certain groups, and even identify and locate individuals.[43]

Enabling discriminatory profiling: Facial recognition software is not just being used to surveil and identify, but also to target and discriminate.[44]

Assisting the spread of disinformation: AI can be used to create and disseminate targeted propaganda, and that problem is compounded by AI-powered social media algorithms driven by "engagement," which promote content most likely to be clicked on. Machine learning powers the data analysis social media companies use to create profiles of users for targeted advertising. In addition, bots disguised as real users further spread content outside of narrowly targeted social media circles by both sharing links to false sources and actively interacting with users as chatbots using natural language processing.[45]

In addition, the specter of "deep fakes," AI systems capable of creating realistic-sounding video and audio recordings of real people, is causing many to believe the technology will be used in the future to create forged videos of world leaders for malicious ends. Although it appears that deep fakes have yet to be used as part of real propaganda or disinformation campaigns, and the forged audio and video is still not good enough to seem completely human, the AI behind deep fakes continues to advance, and there is potential for sowing chaos, instigating conflict, and further causing a crisis of truth that should not be discounted.[46]

Perpetuating bias in the job market: Hiring processes have long been fraught with bias and discrimination. In response, an entire industry has emerged that uses AI with the goal of removing human bias from the process. However, many products ultimately risk perpetuating the very bias they seek to mitigate. As in other areas a major cause of this is the prevalent use of historical data of past "successful" employees to train the ML models, thus naturally reproducing the bias in prior hiring.[47]

Driving financial discrimination against the marginalized: Algorithms have long been used to create credit scores and inform loan screening. However, with the rise of big data, systems are now using machine learning to incorporate and analyze non-financial data points to determine creditworthiness, from where people live, to their internet browsing habits, to their purchasing decisions. The outputs these systems produce are known as e-scores, and unlike formal credit scores they are largely unregulated. As data scientist Cathy O'Neil has pointed out, these scores are often discriminatory and create pernicious feedback loops.[48]

43 Recently, Amazon has come under fire for directly marketing a facial recognition product called Rekognition to law enforcement agencies for use in conjunction with police body cameras, which would allow police to identify people in real time. The product was piloted with police departments in Orlando, Florida and Washington County, Oregon. https://www.theguardian.com/technology/2018/may/22/amazon-rekognition-facial-recognition-police

44 One example is an Israeli company called Faception, which bills itself as a "facial personality analytics technology company," and claims it can categorize people into personality types based solely on their faces. The classifiers it uses include "white collar offender," "high IQ," "paedophile" and "terrorist." The company has not released any information about how its technology can correctly label people based only on their faces. See: Paul Lewis, "'I was shocked it was so easy': meet the professor who says facial recognition can tell if you're gay," The Guardian, July 7, 2018.

45 Given bots are estimated to make up at least half of all internet traffic, their reach should not be underestimated. See: Michael Horowitz, Paul Scharre, Gregory C. Allen, Kara Frederick, Anthony Cho and Edoardo Saravalle, "Artificial Intelligence and International Security," Center for a New American Security, July 10, 2018, https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security.

46 Ibid.

47 Monica Torres, "Companies Are Using AI to Screen Candidates Now with HireVue," Ladders, August 25, 2017, https://www.theladders.com/career-advice/ai-screen-candidates-hirevue.

48 For example, a would-be borrower who lives in a rough part of town, where more people default on their loans, may be given a low score and targeted with financial products offering less credit and higher interest rates. This is because such systems group people together based on the observed habits of the majority. In this case, a responsible person trying to start a business could be denied credit or given a loan on unfavorable terms, perpetuating existing bias and social inequality. O'Neil, 141-160. One company O'Neil singled out is ZestFinance, which uses machine learning to offer payday loans at lower rates than typical payday lenders. The company's philosophy is "all data is credit data." Some of the data has been found to be a proxy for race, class, and national origin. This includes whether applicants use proper spelling and capitalization on their application, and how long it takes them to read it. Punctuation and spelling mistakes are analyzed to suggest the applicant has less education and/or is not a native English speaker, which are highly correlated with socioeconomic status, race, and national origin. This means those who are considered to have poor language skills -- including non-native speakers -- will have higher interest rates. This can lead to a feedback loop that entrenches existing discriminatory lending practices -- if the applicants have trouble paying these higher fees, this tells the system that they were indeed higher risk, which will result in lower scores for other similar applicants in the future. O'Neil, 157-158.

VI. AI AND HUMAN RIGHTS

WHY DO HUMAN RIGHTS MATTER?

AI has "created new forms of oppression, and in many cases disproportionately affects the most powerless and vulnerable. The concept of human rights addresses power differentials and provides individuals, and the organizations that represent them, with the language and procedures to contest the actions of more powerful actors, such as states and corporations."[49]

Human rights are universal and binding, and are codified in a body of international law. Respecting human rights is required of both governments and companies alike, although governments have additional obligations to protect and fulfill human rights.[50] There is an entire system of regional, international, and domestic institutions and organizations that provide well-developed frameworks for remedy and articulate the application of human rights law to changing circumstances, including technological developments. And in cases where domestic law is lacking, the moral legitimacy of human rights carries significant normative power.[51] Violating human rights carries global reputational and political costs, and naming and shaming human rights violators is often an effective tool. Human rights law can address some of the most egregious societal harms caused by AI, and prevent such harms from occuring in the future.

Ethics and its role as a complementary area: *Until now, the ethics discourse has largely dominated the discussion about "good" and "bad" AI. This focus is understandable, and ethics do play an important role. Artificial intelligence has sparked more discussions about the interplay between human beings and machines than perhaps any previous technological development. Considering ethical concepts such as justice, fairness, transparency, and accountability allows for valuable debate about the societal impacts of AI, and the role of AI in our lives.*[52] *There is also an academic research community devoted to addressing ethical issues.*[53] *Ethics have helped those researching and developing AI to define boundaries for themselves. Major AI players such as Google, Microsoft, and DeepMind have developed ethical principles to guide how they pursue AI initiatives.*[54]

Human rights are more universal and well-defined than ethics principles, and they provide for accountability and redress. In this way, human rights and ethics can be mutually reinforcing. For example, a company may develop ethical AI principles such as avoiding reinforcing negative social biases and making sure their systems are accountable to human oversight. The human rights of privacy and non-discrimination, among others, can then further define those ethical principles, and the international human rights regime can provide for remedy should those principles be violated. Additionally, if a use of AI is deemed unethical, it is likely that it also violates human rights, and the principles and procedures embedded in the international human rights regime can be leveraged to combat that unethical use of AI. We discuss recommendations for how stakeholders can use both ethics and human rights together internal policies below.

49 https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5
50 According to the UN Principles on Business and Human Rights, States must protect against human rights abuse by businesses within their jurisdiction, businesses are responsible for respecting human rights wherever they operate, victims must have access to judicial and non judicial remedy. For more information, see: https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf
51 Ibid.
52 https://www.considerati.com/publications/blog/marrying-ethics-human-rights-ai-scrutiny/
53 The Fairness, Accountability and Transparency in Machine Learning initiative: https://www.fatml.org/
54 For each policy see -- Microsoft: https://www.microsoft.com/en-us/ai/our-approach-to-ai, Google: https://www.blog.google/technology/ai/ai-principles/, DeepMind: https://deepmind.com/applied/deepmind-ethics-society/principles/

HOW AI IMPACTS HUMAN RIGHTS

The role of AI in facilitating discrimination is well documented, and is one of the key issues in the ethics debate today. To recognize these issues, Access Now partnered with human rights organizations and AI companies to release "The Toronto Declaration" in March 2018.[55] However, the right to non-discrimination is not the only human right implicated by AI. Because human rights are interdependent and interrelated, AI affects nearly every internationally recognized human right.

Below we examine many of the human rights impacted by AI.[56] The rights discussed are largely those embodied in the three documents that form the base of international human rights law, the so-called "International Bill of Human Rights."[57] This includes the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR), and the International Covenant on Economic, Social and Cultural Rights (ICESCR).[58] To these, this report adds the right to data protection as defined by the EU Charter of Fundamental Rights.[59] For each implicated human right we discuss how current AI uses violate or risk violating that right, as well as risks posed by prospective future developments in AI. It is important to note that the human rights issues discussed below are not necessarily unique to AI. Many already exist within the digital rights space, but the ability of AI to identify, classify, and discriminate magnifies the potential for human rights abuses in both scale and scope.

Like the human rights harms in other uses of technology that leverage data, the harms related to the use of AI often disproportionately impact marginalized populations.[60] That can include women and children, as well as certain ethnic, racial, or religious groups, the poor, the differently abled, and members of the LGBTQ community. The long-established marginalization of these groups is reflected in the data and reproduced in outputs that entrench historic patterns.

Rights to life, liberty and security, equality before the courts, a fair trial[61]

> "*Everyone has the right to liberty and security of person. No one shall be subjected to arbitrary arrest or detention. No one shall be deprived of his liberty except on such grounds and in accordance with such procedure as are established by law.*" - Article 9 of the ICCPR

> "*All persons shall be equal before the courts and tribunals. In the determination of any criminal charge against him, or of his rights and obligations in a suit at law, everyone shall be entitled to a fair and public hearing by a competent, independent and impartial tribunal established by law [...] Everyone charged with a criminal offense shall have the right to be presumed innocent until proven guilty according to law.*" - Article 14 of the ICCPR

55 "The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems, https://www.accessnow.org/cms/assets/uploads/2018/05/Toronto-Declaration-D0V2.pdf.
56 Human Rights not included are: freedom from torture, right not to be enslaved, rights of detainees, right not to be imprisoned merely based on inability to fulfill a contractual obligation, rights of aliens, and the right to social security. This does not mean that AI cannot ultimately impact these rights, merely that we found no current documented violations, nor and prospective violations we believed could occur in the near future.
57 Note that although there are many regional human rights systems that are more comprehensive, we mostly limited our analysis to the UN-based system in the interest of universal applicability. The exception to this is the right to data protection, which Access Now recognizes as a right and is particularly relevant in the context of AI. Further analysis of AI relating to the rights enumerated in these regional systems is merited. For example, the European Convention on Human Rights and the EU Charter of Fundamental Rights are far more comprehensive when it comes to workers' rights, and use of AI by employers to monitor and police employee activity may violate European human rights.
58 See https://www.ohchr.org/EN/ProfessionalInterest/Pages/InternationalLaw.aspx for more information
59 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN
60 For a more in-depth examination of how this plays out in the U.S., see Automating Inequality by Virginia Eubanks. See also https://harpers.org/archive/2018/01/the-digital-poorhouse/ ("the most marginalized in our society face higher levels of data collection when they access public benefits, walk through heavily policed neighborhoods, enter the health care system, or cross national borders. That data reinforces their marginality when it is used to target them for extra scrutiny. Groups seen as undeserving of social support and political inclusion are singled out for punitive public policy and more intense surveillance, and the cycle begins again. It is a feedback loop of injustice.").
61 Article 3, 6, 7, 8, and 10 of UDHR, Articles 9 and 14 of the ICCPR

> *"Every human being has the inherent right to life. This right shall be protected by law. No one shall be arbitrarily deprived of his life. In countries which have not abolished the death penalty, sentence of death may be imposed only for the most serious crimes in accordance with the law in force at the time of the commission of the crime and not contrary to the provisions of the present Covenant."* - Article 6 of the ICCPR

The growing use of AI in the criminal justice system risks interfering with rights to be free from interferences with personal liberty. One example is in recidivism risk-scoring software used across the U.S. criminal justice system to inform detainment decisions at nearly every stage, from assigning bail to criminal sentencing.[62] The software has led to more black defendants falsely labeled as high risk and given higher bail conditions, kept in pre-trial detention, and sentenced to longer prison terms. Additionally, because risk-scoring systems are not prescribed by law and use inputs that may be arbitrary, detention decisions informed by these systems may be unlawful or arbitrary.

Criminal risk assessment software is pegged as a tool to merely assist judges in their sentencing decisions. However, by rating a defendant as high or low risk of reoffending, they attribute a level of future guilt, which may interfere with the presumption of innocence required in a fair trial.[63] Predictive policing software also risks wrongly imputing guilt, building in existing police bias through the use of past data. Reports suggest that judges know very little about how such risk-scoring systems work, yet many rely heavily upon the results because the software is viewed as unbiased.[64] This raises the question of whether or not court decisions made on the basis of such software can truly be considered fair.[65]

When they use these tools, governments essentially hand over decision making to private vendors. The engineers at these vendors, who are not elected officials, use data analytics and design choices to code policy choices often unseen by both the government agency and the public. When individuals are denied parole or given a certain sentence for reasons they will never know and that cannot be articulated by the government authority charged with making that decision, trials may not be fair and this right may be violated.[66]

Looking forward: Broadly deployed, facial recognition software within law enforcement raises the risk of unlawful arrest due to error and overreach. History is rife with examples of humans wrongly arresting people who happen to look similar to wanted criminals.[67] Given the error rates of current facial recognition technology, these inaccuracies could lead to increased wrongful arrests due to misidentification, exacerbated by the lower accuracy rates for non-white faces.[68]

The inability of AI to deal with nuance will likely cause more problems in the future. Laws are not absolute; there are certain cases where breaking the law is justified. For example, it is probably acceptable to run a red light in order to avoid a rear-end collision with a tailgating car. While a human police officer can make that distinction, and elect not to ticket the driver, red light cameras are not capable of such judgment. In a future of AI-powered smart cities and "robocops," there is a risk that this loss of nuance will lead to a drastic increase in people wrongfully arrested, ticketed, or fined, with limited recourse. Over time these circumstances could push us into a world where people preference strictly following any law or rule despite extenuating circumstances, losing the ability to make necessary judgment calls.

62 Angwin et. al, "Machine Bias."
63 According to the General Comment 32 on Article 14 of the ICCPR
64 Angwin et. al, "Machine Bias."
65 As was previously discussed, not all communities are policed equally, and because of this bias AI-powered software ultimately creates negative feedback loops that can "predict" increasing criminal activity in certain areas, resulting in continually overpoliced communities. See: The Committee of Experts on Internet Intermediaries, "Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications," Council of Europe, March 2018, pg. 10-12, https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5.
66 Robert Brauneis and Ellen P. Goodman, "Algorithmic Transparency for the Smart City," The Yale Journal of Law and Technology, Vol. 20 (2018), 103-176, https://www.yjolt.org/sites/default/files/20_yale_j._l._tech._103.pdf.
67 "Face Value," IRL: Online Life is Real Life, Podcast audio, February 4, 2018, https://irlpodcast.org/season2/episode3/.
68 Lauren Goode, "Facial recognition software is biased towards white men, researcher finds," the Verge,Feb. 11, 2018, https://www.theverge.com/2018/2/11/17001218/facial-recognition-software-accuracy-technology-mit-white-men-black-women-error.

With the availability of increasingly more data about our lives, it is foreseeable that information such as social media posts and activity will be included in AI-based systems that inform law enforcement and judicial decisions. ML could be harnessed to identify language or behaviors that show a propensity for violence, or a risk of committing certain types of crimes. Such a use would further implicate the rights to equality under the law and a fair trial.

Rights to privacy and data protection[69]

> "*No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.*" - Article 17 of the ICCPR

> "*Everyone has the right to respect for his or her private and family life, home and communications.*" - Article 7 of the EU Charter of Fundamental Rights

> *"Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified."* - Article 8 of the EU Charter of Fundamental Rights

Privacy is a fundamental right that is essential to human dignity. The right to privacy also reinforces other rights, such as the rights to freedom of expression and association.[70] Many governments and regions now recognize a fundamental right to data protection. Data protection is primarily about protecting any personal data related to you.[71] It is closely related to the right to privacy, and can even be considered a part of the right to privacy within the UN human rights system.

AI systems are often trained through access to and analysis of big data sets. Data are also collected in order to create feedback mechanisms and provide for calibration and continual refinement. This collection of data interferes with rights to privacy and data protection. The analysis of data using AI systems may reveal private information about individuals, information that qualifies as protected information and should be treated as sensitive even if derived from big data sets fed from publicly available information. For example, researchers have developed ML models that can accurately estimate a person's age, gender, occupation, and marital status just from their cell phone location data. They were also able to predict a person's future location from past history and the location data of friends.[72] In order to protect human rights, this information must be treated the same as any other personal data.

Another example of the thin line between public and private data is the increased use of government social media monitoring programs, wherein law enforcement agencies collect troves of social media information and feed it to AI-powered programs to detect alleged threats. While isolated checks of a target's public social media may seem to some like a wise policing strategy, these programs instead will involve massive, unwarranted intake of the entire social media lifespan of an account, group of accounts, or more. Bulk collection of this type has been found to inherently violate human rights. Additionally, if the systems used

69 Article 12 of UDHR, Article 17 of ICCPR, Article 8 of the EU Charter of Fundamental Rights
70 "Necessary and Proportionate Principles"
71 Estelle Masse, "Data Protection: why it matters and how to protect it," Access Now, January 25, 2018, https://www.accessnow.org/data-protection-matters-protect/.
72 Steven M. Bellovin, et. al, "When enough is enough: Location tracking, mosaic theory, and machine learning," NYU Journal of Law and Liberty, 8(2) (2014) 555--628, https://digitalcommons.law.umaryland.edu/fac_pubs/1375/

to process data are insufficiently transparent or accountable such that it's unclear in human terms how decisions are reached, the systems violate key elements of the right to data protection.

Looking forward: The risks due to ability of AI to track and analyze our digital lives are compounded because of the sheer amount of data we produce today as we use the internet. With the increased use of Internet of Things (IoT) devices and the attempts to shift toward "smart cities," people will soon be creating a trail of data for nearly every aspect of their lives. Although the individual pieces of this data may seem innocuous, when aggregated they reveal minute details about our lives. AI will be used to process and analyze all this data for everything from micro-targeted advertising, to optimizing public transportation, to government surveillance of citizens. In such a world, not only are there huge risks to privacy, but the situation raises the question of whether data protection will even be possible.

Government surveillance has expanded with the growth of the internet and the development of new technologies, and AI is enabling more invasive surveillance tools than ever. For example, although no fully centralized government facial recognition system is yet known to exist, China's work toward installing more CCTV cameras in public places and centralizing its facial recognition systems shows that this could soon change. In the U.S., half of all adults are already in law enforcement facial recognition databases.[73] Their use threatens to end anonymity, and the fear of being watched can stop people from exercising other rights, such as the freedom of association. The negative impact of AI-powered surveillance would be felt most acutely by the marginalized populations who are disproportionately targeted by security forces.[74] Additionally, because 24/7 monitoring of the general population is neither necessary nor proportionate to the goal of public safety or crime prevention,[75] it would almost certainly violate the fundamental right to privacy.

Right to freedom of movement

> *"Everyone lawfully within the territory of a State shall, within that territory, have the right to liberty of movement and freedom to choose his residence. Everyone shall be free to leave any country, including his own. The above-mentioned rights shall not be subject to any restrictions except those which are provided by law, are necessary to protect national security, public order (ordre public), public health or morals or the rights and freedoms of others, and are consistent with the other rights recognized in the present Covenant. No one shall be arbitrarily deprived of the right to enter his own country."* - Article 12 of the ICCPR

The potential for AI to restrict freedom of movement is directly tied to its use for surveillance. In systems that combine data from satellite imagery, facial recognition-powered cameras, and cell phone location information, among other things, AI can provide a detailed picture of individuals' movements as well as predict future location. It could therefore easily be used by governments to facilitate more precise restriction of the freedom of movement, at both the individual and group level.

Looking forward: Currently, the lack of formal mapping in many poor and underserved communities around the world has led to exclusion from GPS mapping apps. Given the growing trend of AI for predictive policing, it is possible that increased mapping of these areas and combining use of that information with data from law enforcement apps, such as those that rate the crime levels and safety of neighborhoods, could effectively shut down tourism or inhibit movement around or within an area. Even if this is done for legitimate public safety reasons, this may risk violating freedom of movement.

73 Jordan G. Telcher, "What Do Facial Recognition Technologies Mean for Our Privacy?" The New York Times, July 18, 2018, https://www.nytimes.com/2018/07/18/lens/what-do-facial-recognition-technologies-mean-for-our-privacy.html?nytapp=true&smid=nytcore-ios-share.
74 Evan Selinger, "Amazon Needs to Stop Providing Facial REcognition Tech for the Government," Medium, June 21, 2018, https://medium.com/s/story/amazon-needs-to-stop-providing-facial-recognition-tech-for-the-government-795741a016a
75 See"The Necessary and Proportionate Principles," and Privacy International, "Guide to International Law and Surveillance," August 2017, https://privacyinternational.org/sites/default/files/2017-12/Guide%20to%20International%20Law%20and%20Surveillance%20August%202017.pdf.

As IoT extends to infrastructure and transportation systems, from smart highways to biometrically tagged public transportation systems, AI will continue to be used to locate individuals in real time, allowing governments to further restrict freedom of movement. Additionally, if AI is used to automate decisions about who can travel—for example, by placing people on a "Do Not Fly" or other prohibitive travel list—errors could result in people having their freedom of movement unjustly restricted.

Rights to freedom of expression, thought, religion, assembly, and association[76]

"Everyone shall have the right to freedom of thought, conscience and religion. This right shall include freedom to have or to adopt a religion or belief of his choice, and freedom, either individually or in community with others and in public or private, to manifest his religion or belief in worship, observance, practice and teaching. No one shall be subject to coercion which would impair his freedom to have or to adopt a religion or belief of his choice." - Article 18 of ICCPR and Article 18 of UDHR

"*Everyone shall have the right to hold opinions without interference. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."* - Article 19 of the ICCPR

"*The right of peaceful assembly shall be recognized. [...] Everyone shall have the right to freedom of association with others, including the right to form and join trade unions for the protection of his interests. No restrictions may be placed on the exercise of this right other than those which are prescribed by law and which are necessary in a democratic society in the interests of national security or public safety, public order (ordre public), the protection of public health or morals or the protection of the rights and freedoms of others."* - Articles 21 and 22 of the ICCPR

Direct Threats: Internet companies that host content use AI to flag posts that violate their terms of service. Governments exerting formal and informal pressure on companies to address the problem of alleged terrorist content, hate speech, and so-called "fake news," but without clear standards or definitions, has led to increased use of automated systems.[77] A law recently passed in Germany requires social media sites to remove a wide range of content within 24 hours after it has been flagged (or up to seven days in cases that are less clear-cut).[78] Because AI is imperfect and companies are pressured to take down questionable content so quickly, much of the content is removed in error.[79] YouTube removed more than 100,000 videos documenting atrocities in Syria after they were flagged, . These videos often serve as the only evidence of horrific crimes and human rights violations, and YouTube's policy carves out exceptions for violent content when it is of important educational or documentary value. [80] Yet they were still taken down.

Authoritarian governments can use similar technology to increase censorship. The Chinese government is already replacing some of its human censors with AI. Popular Chinese video platform iQiyi uses ML to identify pornographic and violent content, as well as content deemed "politically sensitive." Because ML cannot deal with nuance, flagged content is currently reviewed by humans, though this may change as the technology becomes more sophisticated and industry sees the human resources required for review as an unnecessary expense.[81]

76 Article 19 of the UDHR and Article 19 of the ICCPR; Article 18 of the ICCPR and UDHR, Articles 21 and 22 of the ICCPR, Article 20 of the UDHR
77 A Freedom House survey found 30 of 65 of governments attempted to control online discussions. https://freedomhouse.org/article/new-report-freedom-net-2017-manipulating-social-media-undermine-democracy.
78 "Germany starts enforcing hate speech law," The BBC, January 1, 2018, https://www.bbc.com/news/technology-42510868.
79 Denis Nolasco and Peter Micek, "Access Now responds to Special Rappoteur Kaye on 'Content Regulation in the Digital Age, Access Now, January 11, 2018
80 Kate O'Flaherty, "YouTube keeps deleting evidence of Syrian chemical weapon attacks," Wired, June 26, 2018, http://www.wired.co.uk/article/chemical-weapons-in-syria-youtube-algorithm-delete-video.
81 Yuan Tang, "Artificial intellifence takes jobs from Chinese censors," Financial Times, May 21, 2018, https://www.ft.com/content/9728b178-59b4-11e8-bdb7-f6677d2e1ce8.

In countries where freedom of religion is under threat, AI could assist government officials in monitoring and targeting members of persecuted religious groups. Not only could this force such groups further into secrecy for fear of being identified, but it could produce physical consequences, from violence to arrest to death. AI could also be used to identify and take down religious content. This would constitute a direct violation of freedom of religion if people are not able to display religious symbols, pray, or teach about their religion online.

Finally, AI-enabled censorship can be used to restrict the freedom of association by removing groups, pages, and content that facilitate organization of in-person gatherings and collaboration. Given the important role of social media in organizing protest movements globally, use of AI could have the widespread effect of hindering assembly worldwide.[82]

Indirect Threats: Violations of the right to privacy have a chilling effect on free expression. When people feel that they are being watched, or lack anonymity, they have been shown to self-censor and alter their behavior. AI-powered surveillance only compounds this effect, which will have serious repercussions for freedom of expression.[83] One powerful example is facial recognition. If used in public spaces to identify individuals at a protest, this may have a significant chilling effect on assembly. The implementation of such a system in countries that restrict free assembly would effectively prevent enjoyment of this right, as many people rely on the level of security anonymity provides to gather in public and express their views.

Another indirect threat is the impact of AI-powered social media and search algorithms. For example, Facebook's algorithm determines the content of a user's newsfeed and influences how widely and to whom content is shared. Google's search algorithm indexes content, and decides what shows up in the top of search results. These algorithms have played a significant role in establishing and reinforcing echo chambers, and they ultimately risk negative impacts on media pluralism and inhibition of a diversity of views.[84]

The role of AI in content ranking and the creation and reinforcement of filter bubbles poses an indirect threat to freedom of thought because it shapes the type of information people have access to. Although people often have the ability to access other sources of information or seek out different opinions, humans' limited time and attention mean most people do not do this. And in countries without a robust free press and limited internet access, social media platforms such as Facebook are often the only source of unregulated information.

Looking forward: A looming direct threat to free expression is through bot-enabled online harassment. While harassment is not new, it is increasingly perpetrated by bots instead of humans. These bot accounts masquerade as real users and send automated responses to identified accounts or to anyone who shares a certain opinion.[85] This kind of relentless online harassment has a chilling effect on free expression, particularly for those in marginalized populations, who are disproportionately targeted.[86] As bot designers increasingly employ natural language processing, harassment bots will follow suit. This will make it harder to detect, report, and get rid of bot accounts.

The predictive power of AI is already used to predict and help prevent armed conflict. This same approach could also be used pre-emptively by governments to predict and prevent public demonstrations or protests before they take place.[87]

82 Alex Comninos, "Freedom of Peaceful Assembly and Freedom of Association and the Internet," APC, https://www.apc.org/sites/default/files/cyr_english_alex_comninos_pdf.pdf.
83 Privacy International and Article 19, "Privacy and Freedom of Expression in the Agen of Artificial Intelligence," April 2018, https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf.
84 Council of Europe, "Algorithms and Human Rights."
85 Michael Bernstein, "Identifying Harassment Bots on Twitter," Daemo, August 17, 2017, https://www.daemo.org/demo/botcheck
86 Megan White, "How do you solve a problem like troll armies?" Access Now, April 21, 2017, https://www.accessnow.org/solve-problem-like-troll-armies/, and Constance Grady, "Obline harassment threatens free speech. Now there's a field guide to help survive it," Vox, May 2, 2018, https://www.vox.com/culture/2018/5/2/17292258/pen-america-online-harassment-field-manual-take-back-the-net
87 Council of Europe, "Algorithms and Human Rights."

Rights to equality and non-discrimination[88]

"All persons are equal before the law and are entitled without any discrimination to the equal protection of the law. In this respect, the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." - Article 26 of the ICCPR

"In those States in which ethnic, religious or linguistic minorities exist, persons belonging to such minorities shall not be denied the right, in community with the other members of their group, to enjoy their own culture, to profess and practise their own religion, or to use their own language." - Article 27 of the ICCPR

"The States Parties to the present Covenant undertake to ensure the equal right of men and women to the enjoyment of all [...] rights set forth in the present Covenant." - Article 3 of the ICCPR and the ICESCR

AI models are designed to sort and filter, whether by ranking search results or categorizing people into buckets. This discrimination can interfere with human rights when it treats different groups of people differently. Sometimes such discrimination has positive social aims, for example, when it is used in programs to promote diversity. In criminal justice, this discrimination is often the result of forms of bias. Use of AI in some systems can perpetuate historical injustice in everything from prison sentencing to loan applications.

Although people may not think online advertisements have much of an impact on their lives, research suggests the online ad space can result in discrimination and perpetuate historical biases. In 2013, researcher Latanya Sweeney found that a Google search for stereotypically African American-sounding names yielded ads that suggested an arrest record (such as "Trevon Jones, Arrested?") in the vast majority of cases.[89] In 2015, researchers at Carnegie Mellon found Google displayed far fewer ads for high-paying executive jobs to women. Google's personalized ad algorithms are powered by AI, and they are taught to learn from user behavior. The more people click, search, and use the internet in racist or sexist ways, the more the algorithm translates that into ads. This is compounded by discriminatory advertiser preferences, and becomes part of a cycle. "How people perceive things affects the search results, which affect how people perceive things."[90]

Looking forward: Given that facial recognition software has higher error rates for darker-skinned faces, it is likely that misidentification will disproportionately affect people of color. The gravity of the problem is demonstrated by the ACLU's test of Amazon's Rekognition facial recognition software. The ACLU scanned the faces of all 535 U.S. members of Congress against 25,000 public criminal mugshots using Rekognition's API with the default 80% confidence level. No one in the U.S. Congress is actually in the mugshot database, yet there were 28 false matches. Of these matches, 38% were people of color, even though only 20% of members of Congress are people of color. [91]

AI-powered surveillance software can also be used with the express purpose of discrimination, allowing governments to identify, target, and deny services to people from different groups. In 2017, a controversial study found that a ML system could accurately guess whether someone was gay or straight, supposedly based solely on photos of their faces. Other experts strongly refuted the findings, pointing out that there are numerous non-facial cues that the ML could have picked up on in the photos. However, regardless

88 Articles 3, 26 and 27 of the ICCPR. Article 3 of the ICESCR
89 Latanya Sweeney, "Discrimination in Online Ad Delivery," Harvard University, January 28, 2013, https://arxiv.org/ftp/arxiv/papers/1301/1301.6822.pdf.
90 Julia Carpenter, "Google's algorithm shows prestigious job ads to men, but not to women. Here's why that should worry you." The Washington Post, July 6, 2015, https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/?noredirect=on&utm_term=.a5cbea41ad6b.
91 This showed that as with many facial recognitions systems, Rekognition disportionately impacted people of color. See: Russell Brandom, "Amazon's facial recognition matched 28 members of Congress to criminal mugshots," The Verge, July 26, 2018, https://www.theverge.com/2018/7/26/17615634/amazon-rekognition-aclu-mug-shot-congress-facial-recognition

of the quality of the study, the model was able to identify the sexuality of 81% of men and 74% of women with accuracy. Governments could use systems like this to target and discriminate against LGBTQ people in places where homosexuality and gender nonconformity is either illegal or social unacceptable. The questionable science behind the study of faces, or the high error rates for this kind of system, may not matter to those wielding the technology. [92]

Rights to political participation and self determination[93]

"Every citizen shall have the right and the opportunity [...] to take part in the conduct of public affairs, directly or through freely chosen representatives; to vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors; to have access, on general terms of equality, to public service in his country." - Article 25 of the ICCPR

The role of AI in creating and spreading disinformation challenges the notion of fair elections and creates a threat to the right to political participation and self determination. The 2016 U.S. presidential election showed how a foreign power can leverage bots and social media algorithms to increase the reach of false information and potentially influence voters. Although platforms are working to prevent this type of activity, a future of AI-powered chatbots and deep fakes will likely make such content more convincing to voters and harder for companies to detect. This may chill political participation, particularly if voters lose trust in the legitimacy of elections.

Looking forward: AI-powered surveillance could be used to restrict and inhibit political participation, including by identifying and discouraging certain groups of people from voting. Use of facial recognition in polling places or voting booths could compromise the secrecy of the ballot. Governments wishing to discourage voters from casting ballots for the opposition need not even directly surveil the act of voting; the mere signification of surveillance could be sufficient to convince voters that their ballots are not secret, and could influence their voting decisions accordingly.

Prohibition on propaganda[94]

"Any propaganda for war shall be prohibited by law. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law." - Article 20 of the ICCPR

Looking forward: Just as people can use AI-powered technology to facilitate the spread of disinformation or influence public debate, they can use it to create and propagate content designed to incite war, discrimination, hostility, or violence. We can see the potential for this scenario in the disputes between Russia and Ukraine over Crimea. Russia has tested now widely known disinformation tactics in attempts to undermine the public's faith in the media.[95] Governments around the world have deployed "troll armies" to stoke the flames of conflict for political ends.[96] In the near future, they could use chatbots to incite racial and ethnic violence in regions that are already rife with tension, or deploy deep fakes to simulate world leaders declaring war or instigating armed conflict.

92 Sam Levin, "New AI can guess whether you're gay or straight from a photograph," The Guardian, September 7, 2017, https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph, and Lewis, "Facial recognition can tell if you're gay." https://www.theguardian.com/technology/2018/jul/07/artificial-intelligence-can-tell-your-sexuality-politics-surveillance-paul-lewis?CMP=Share_iOSApp_Other.

93 Article 21 of the UDHR, Article 25 of the ICCPR

94 Article 20 of the ICCPR

95 After the massive 2014 pro-democracy protests in Ukraine that resulted in the ouster of Pro-Russian president Viktor Yanukovych, Russia began a massive disinformation campaign to discredit the new government and encourage separatists to initiate the current civil war. See: Gregory Warner, "What Americans Can Lean From Fake News in Ukraine," Rough Translation, audio podcast, August 21, 2017, https://www.npr.org/2017/08/21/544952989/rough-translation-what-americans-can-learn-from-fake-news-in-ukraine.

96 White, "troll armies," Access Now.

Rights to work, an adequate standard of living[97]

"The States Parties to the present Covenant recognize the right to work, which includes the right of everyone to the opportunity to gain his living by work which he freely chooses or accepts, and will take appropriate steps to safeguard this right. The steps to be taken by a State Party to the present Covenant to achieve the full realization of this right shall include technical and vocational guidance and training programmes, policies and techniques to achieve steady economic, social and cultural development and full and productive employment under conditions safeguarding fundamental political and economic freedoms to the individual." - Article 6 of the ICESCR

"The States Parties to the present Covenant recognize the right of everyone to an adequate standard of living for himself and his family, including adequate food, clothing and housing, and to the continuous improvement of living conditions." - Article 11 of the ICESCR

Although the right to work does not constitute the absolute and unconditional right to obtain employment, it does require states to work toward achieving full employment.[98] The role of AI in the automation of jobs could poses a real threat to the right to work; it may prevent some people from accessing the labor market in the first place. Automation has resulted in job loss in certain sectors, and AI is widely predicted to accelerate this trend. Although there is significant disagreement as to the extent that job automation will be achieved, there is no doubt that AI will result in some shifts in the labor market, both through job creation and job destruction.[99]

Looking forward: If automation does shift the labor market significantly, and large numbers of people cannot find jobs, they will struggle to provide for themselves and their families. Researchers are exploring ways to ensure people can maintain an adequate standard of living with volatility in the labor market. One approach is a universal basic income, a fixed income that governments provide. Canada, Finland, and California are all testing out basic income schemes, and more trials are planned in other countries. [100]

Job automation may bring about a range of challenges that governments will have to address to ensure an adequate standard of living. In the U.S., the government uses automated decision-making systems in programs to address poverty, for everything from eligibility for government-funded health care to food assistance.[101] During his 2017 visit to the U.S., the UN Special Rapporteur on extreme poverty and human rights found that city authorities across the country are using automated systems to match the homeless population with available services. These systems use traditional deterministic statistical algorithms to assign a homeless respondent a "vulnerability score" and then connect the person to appropriate housing opportunities.[102] The existence of such systems raises important questions about automating these crucial decisions, but at the very least, they produce traceable outcomes. However, if there is a shift to using ML, the inherent lack of transparency and explainability of ML could make automated decisions about the provision of public service something that neither the government agency tasked with making the decision nor the public fully understands.

97 Articles 23 and 25 of the UDHR, Articles 6, 7, 11 of the ICESCR
98 See http://hrlibrary.umn.edu/gencomm/escgencom18.html
99 See https://www.technologyreview.com/s/610005/every-study-we-could-find-on-what-automation-will-do-to-jobs-in-one-chart/.
100 Leonid Bershidsky, "Finland's Basic Income Test Wasn't Ambitious Enough," Bloomberg Opinion, April 26, 2018, https://www.bloomberg.com/view/articles/2018-04-26/finland-s-basic-income-experiment-was-doomed-from-the-start, Chis Weller, "One of the biggest VCs in Silicon Walley is launching an experiment that will gie 3,000 people free money until 2022," Business Insider, September 21, 2017,https://www.businessinsider.com/y-combinator-basic-income-test-2017-9, and Jordan Pearson, "Basic Income Is Already Transforming Life and Work In a Postindustrial Canadian City," Motherboard, April 23, 2018, https://motherboard.vice.com/en_us/article/paxzv8/hamilton-canada-basic-income-pilot-future-work-v25n1.
101 Eubanks, "The Digital Poorhouse."
102 "Statement on Visit to the USA, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights," Office of the High Commissioner for Human Rights, United Nations, December 15, 2017, https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=22533&LangID=E.

Right to health

> *"The States Parties to the present Covenant recognize the right of everyone to the enjoyment of the highest attainable standard of physical and mental health. The steps to be taken by the States Parties to the present Covenant to achieve the full realization of this right shall include those necessary for: (a) The provision for the reduction of the stillbirth-rate and of infant mortality and for the healthy development of the child; (b) The improvement of all aspects of environmental and industrial hygiene; (c) The prevention, treatment and control of epidemic, endemic, occupational and other diseases; (d) The creation of conditions which would assure to all medical service and medical attention in the event of sickness."* - Article 12 of the ICESCR

Some of the most promising and impactful applications of AI are in healthcare, from helping doctors more accurately diagnose disease, to providing more individualized patient treatment recommendations, to making specialist medical advice more accessible. However, there are also ways in which AI could endanger the right to health. One is the potential for AI-powered systems to result in discrimination, or be programmed in ways that place outcomes (such as cost reduction) over the wellbeing of the patient.

For example, an AI system could be designed to recommend different treatments depending on the insurance status of the patient or how much they are able to pay, potentially denying lifesaving care to someone because of their socioeconomic status, harming marginalized groups who already suffer from insufficient access to quality healthcare. Another potential issue is the negative feedback loops that could result from over-reliance on the guidance of an AI system. For example, if doctors tend to withdraw care for patients with certain diagnoses, such as extreme premature birth or severe brain injuries, an ML-based system may learn that such diagnoses are nearly always fatal and recommend the doctor not treat, even if in some cases treatment may be effective.[103]

And of course, there is the impact of the inevitable error rates of any system. Even if, for example, IBM's" Watson" is more accurate than human doctors at diagnosing disease, it will still get the diagnosis wrong on occasion, or recommend the wrong treatment. In this case, what kind of accountability is there for a life-and-death medical decision made by a machine vs. a doctor? The same issue could arise in AI systems that predict disease outbreaks and recommend responses. What happens when you deploy resources to an area deemed high risk while leaving others without assistance? Human health workers already make this choice, but AI would do this preemptively, and may sometimes get it wrong. This raises larger questions regarding the extent to which certain things should be automated, how and when to require a "human in the loop," and how much responsibility should be held by human doctors vs. the AI systems for making the recommendations.

Looking forward: Another concern relates to the use of AI to determine who gets access to healthcare and what they pay for health insurance. There is a danger that health insurance providers could use AI for profiling based on certain behaviors and history. An AI system could use data points about you to recommend individualized health insurance rates. It could see you have history of illness in the family, are not physically active, enjoy eating out at fast food restaurants, and smoke, and recommend charging you higher rates based on these factors.

103 Patricia Hannon, "Researchers say use of artificial intelligence in medicine raises ethical questions," Stanford Medicine News Center, March 14, 2018, https://med.stanford.edu/news/all-news/2018/03/researchers-say-use-of-ai-in-medicine-raises-ethical-questions.html.

Right to education[104]

"The States Parties to the present Covenant recognize the right of everyone to education. [...] The States Parties to the present Covenant recognize that, with a view to achieving the full realization of this right: (a) Primary education shall be compulsory and available free to all; b) Secondary education in its different forms, including technical and vocational secondary education, shall be made generally available and accessible to all by every appropriate means, and in particular by the progressive introduction of free education; (c) Higher education shall be made equally accessible to all, on the basis of capacity, by every appropriate means, and in particular by the progressive introduction of free education; (d) Fundamental education shall be encouraged or intensified as far as possible for those persons who have not received or completed the whole period of their primary education; (e) The development of a system of schools at all levels shall be actively pursued, an adequate fellowship system shall be established, and the material conditions of teaching staff shall be continuously improved." - Article 13 of ICESCR

AI can fundamentally violate the principle of equal access. Universities in the U.S. are using deterministic algorithmic systems to recommend applicants they should admit. These are often custom-built to meet the school's preferences, and have a host of issues that can lead to discrimination, including use of historical data of previously admitted students to inform the model. Since many elite universities have historically been attended by wealthy white males, any model that uses these data risks perpetuating past trends.[105] Such systems will likely employ ML in the future, which would make bias harder to detect. This could result in universities discriminating under the guise of objectivity.

Looking forward: If AI is used to track and predict student student performance in such a way that limits the eligibility to study certain subjects or have access to certain educational opportunities, the right to education will be put at risk. Given the growth of research into early childhood predictors of success, it is likely that such a system could be used to restrict the opportunities of students at increasingly younger ages, resulting in significant discrimination, with students coming from underprivileged backgrounds ultimately being denied opportunities because people from that background tend to have more negative outcomes. Such a system would ignore the students that overcome adversity to achieve academic and professional success, and would entrench existing educational inequalities.

Right to take part in cultural life and enjoy benefits of scientific progress[106]

"The States Parties to the present Covenant recognize the right of everyone: (a) To take part in cultural life; (b) To enjoy the benefits of scientific progress and its applications; (c) To benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author." - Article 15 of the ICESCR

The use of AI technologies that allow governments to identify and repress cultural groups could stop people from taking part in cultural life, either directly or indirectly (for example through surveillance that inspires fear of being identified or suffering reprisals for cultural identity, leading people to avoid cultural expressions altogether). There is a risk that AI could be used to "criminalize" certain cultures. When members of a particular culture are disproportionately arrested or otherwise targeted by law enforcement, the behaviors and customs associated with these cultures could become linked with criminal activities. For example, a ML system analyzing video or photographic footage could learn to associate certain types of dress, manners of speaking, or gestures with criminal activity, and could be used to justify the targeting of these groups under the guise of preventing crime.

104 Article 25 of the UDHR. Article 13 and 14 of the ICESCR
105 O'Neil, Weapons of Math Destruction, 50-67.
106 Article 27 of the UDHR, Article 15 of the ICESCR

Many in the developing world worry that they will be "left behind" in the global AI race and the corresponding transformational economic change. But those in developing countries stand to become passive consumers of AI systems developed in China or the West for different people, cultures, and situations. Foreign-developed AI runs the risk of deepening the existing inequality and social division in places where internet access and technology are largely restricted to the wealthy and urban. This risk of deeper inequality is compounded by the risk that job automation will lead to job loss by displacing the manufacturing industry's role in economic development.

Right to marry, children's rights, and family rights[107]

"*The family is the natural and fundamental group unit of society and is entitled to protection by society and the State. The right of men and women of marriageable age to marry and to found a family shall be recognized. No marriage shall be entered into without the free and full consent of the intending spouses.*" - Article 23 of the ICCPR

"*Every child shall have, without any discrimination as to race, colour, sex, language, religion, national or social origin, property or birth, the right to such measures of protection as are required by his status as a minor, on the part of his family, society and the State.*" - Article 24 of the ICCPR

Looking forward: If AI technology is used for health and reproductive screening, and some people are found to be unlikely to have children, screening could prevent them from marrying, or from marrying a certain person if the couple is deemed unlikely to conceive. Similarly, AI-powered DNA and genetics testing could be used in efforts to produce children with only desired qualities.

ROBOTICS AND AI

The use of AI in robotics represents a small percentage of AI use today. However, robotics is a growing field and robots will increasingly play a role in our lives. In many cases, a robot simply provides the physical body for the types of AI systems explored in this report. However, this physicality, and the context in which AI-powered robots are used, may raise new challenges.[108]

Right to life

Fully autonomous weapons systems are currently under development in many countries. The increasing use of drones and similar weaponry mean that autonomous weapons are likely to be accessible to non-state actors that are not bound by traditional laws of armed conflict. Autonomous weapons in the near future are likely to suffer from AI's inability to deal with nuance or unexpected events. In a conflict situation, this could result in the death or injury of innocent civilians that a human operator may have been able to avoid.[109]

Another threat to the right to life could arise from the use of AI-powered robots in healthcare. Robots are now used to assist in surgery, and the existence of fully autonomous surgical robots is easy to imagine in the near future, as are robots that are used for rehabilitative therapy and general care settings. Robots will inevitably get it wrong. What happens when they do? And who is accountable?[110] Additionally, if bad actors interfere with health robots and they are made to cause physical harm, pathways to remedy or redress the harm are far from established.

107 Article 16 of the UDHR, Articles 23 and 24 of the ICCPR, Article 10 of the ICESCR
108 A helpful starting point for thinking about the desired boundaries of robots in society are science fiction author Isaac Asimov's three laws of robotics.# The first two laws have particular relevance for human rights: 1) A robot may not injure a human being or, through inaction, allow a human being to come to harm; 2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. Below we explore some potential threats of AI-powered robots to human rights.
109 OMEST, "Report of COMEST on Robotics Ethics."
110 Ibid.

Right to privacy

Surveillance drones or other robots have long been used by the military, and they are now increasingly used by law enforcement or non-state actors as well. When equipped with AI-powered technology, such as facial recognition technology, and made to be semi- or fully autonomous—for example, used to follow a certain group or a person independently—such drones could deepen the impact of widespread and invasive surveillance that violates the "necessary and proportionate" principles that govern state surveillance.

Right to work

AI-powered robots can enable job automation, and thus they can threaten the right to work in the ways we explore above.

Right to education

Although still in nascent stages, the use of robotics in education is an active research area. This includes robots used for tasks like teaching second languages in primary schools, and for storytelling.[111] As with more general AI, the risks posed by AI-powered robots have to do with outcomes that violate equal access. For example, in areas where robots replace teachers in schools, students would receive a different kind of education than those with human teachers, and that may constitute a violation of equal access.

VII. RECOMMENDATIONS: HOW TO ADDRESS AI-RELATED HUMAN-RIGHTS HARMS

Swift action now to deal with human rights risks can help prevent the foreseeable detrimental impacts of AI, while providing space and a framework for addressing the problems we cannot predict. Because AI is such a large and diverse field, any approach will need to be sector-specific to some extent. However, four broad policy approaches could address many of the human rights risks posed by AI.

1. Comprehensive data protection legislation can anticipate and mitigate many of the human rights risks posed by AI. However, because it is specific to data, additional measures are also necessary.
2. Government use of AI should be governed by a high standard, including open procurement standards, human rights impact assessments, full transparency, and explainability and accountability processes.
3. Given the private sector's duty to respect and uphold human rights, companies should go beyond establishing internal ethics policies and develop transparency, explainability, and accountability processes.
4. Significantly more research should be conducted into the potential human rights harms of AI systems and investment should be made in creating structures to respond to these risks.

THE ROLE OF COMPREHENSIVE DATA PROTECTION LAWS

Comprehensive data protection laws, which should apply to both the government and private sector, can go a long way in addressing many of the human rights risks posed by AI. Because data is the engine of AI, any law that mandates protection of personal data will necessarily implicate AI systems. Given the global push toward data protection legislation, this is both heartening and practical.

Consider the impact of the European Union's General Data Protection Regulation (GDPR). The GDPR is a positive framework that provides for control of a person's personal information and empowers people to make informed decisions about how their data are used. The GDPR limits data processing to permissible

111 Ibid.

purposes, with heightened protections for sensitive data. It also requires opt-in consent,[112] which limits the use of personal data for training AI systems.

Rights provided for by the GDPR, and other similar laws, offer a framework to prevent against unaccountable uses of AI that impact individual rights, while ensuring a level of control of personal data and accountability for the use of AI and ML systems.

Some have suggested that data protection laws are incompatible with AI and we should make broad exceptions for its development and use. That is misguided. While it is likely true that strong data protection laws may preempt deployment of certain AI systems, companies have never been able to "innovate" without regard for potential harm. If AI systems are used to make decisions on a basis or rationale that not even their developers can fully explain, at-risk individuals—or AI "guinea pigs"—will be the first to suffer the negative consequences. Data protection rights not only provide accountability structures to mitigate harm, they also protect people against having their personal data covertly co-opted, commodified, and otherwise exploited in ways that harm others or society at large.

Innovation and AI

Some industries have begun to pick up on these issues and are developing voluntary frameworks for the use of AI.[113] However, history shows that industry self-regulation can be woefully inadequate at protecting people, particularly those in marginalized communities who are frequently targeted by manipulation campaigns. As AI continues to increase in sophistication, society cannot afford to sacrifice individual rights at the altar of innovation. Instead, in jurisdictions without data protection laws, government officials should be pursuing tech-neutral measures to address the human rights impact of the continuing AI revolution. Similarly, in areas with laws already in force, overseers and watchdogs should ensure that the law is followed and remains relevant as AI technology advances.

Data Protection Rights and AI[114]

The Right to Information and **Right to Access** work together to allow people to get information about what data an entity is collecting, how they are collecting it, how they will use it, and whether data will be used for automated decision-making. These rights raise public awareness about the existence of AI systems and the roles they play.[115] Furthermore, these rights allow people to uncover and understand potential human rights harms and push entities to be more transparent about how they use AI.

The Right to Rectification allows people to amend and modify their information held by a third party if it is incorrect, incomplete, or inaccurate. This right can help mitigate the impact of error rates in AI systems.

The Right to Restrict Processing gives people the ability to request that an entity stop using or limit the use of personal information while **the Right to Erasure** provides a pathway for deletion of a person's personal data held by a third party entity when it is no longer necessary, the information has been misused, or the relationship between the user and the entity is terminated. These rights could be used to temporarily halt the use of a contested AI system, or to pressure an entity to use an AI system more responsibly.

112 https://gdpr-info.eu/art-9-gdpr/
113 See, e.g., https://motherboard.vice.com/en_us/article/a34pp4/john-deere-tractor-hacking-big-data-surveillance ("The American Farm Bureau helped construct the "Privacy and Security Principles for Farm Data," which addresses issues of data ownership, portability, use, and sharing. Companies like Deere and Monsanto were early signers, but questions remain about how much these principes protect in practice.").
114 https://www.accessnow.org/cms/assets/uploads/2018/07/GDPR-User-Guide_digital.pdf
115 In the U.S., many of the details of government use of algorithmic decision making system are hidden behind non-disclosure agreements and memorandums of understanding with vendors. See https://www.wired.com/story/when-government-rules-by-software-citizens-are-left-in-the-dark/.

The Right to an Explanation provides for a person to get an explanation about how an automated decision is made pertaining to that person. This right ensures entities understand how the systems they use actually work, and it pushes AI developers to continue working to make AI understandable.

The Right to Object gives people the ability to contest most processing of their personal data by an entity when the data are used for direct marketing, automated decision making (where no human intervention will take place), research and statistics, or for an entity's "legitimate interest." This right allows for direct challenges to decisions made using AI systems. It is particularly important for government use of AI in ways that can be discriminatory. It also ensures there is a human in the loop in important automated decision-making systems, which adds a layer of accountability.

AI-SPECIFIC RECOMMENDATIONS FOR GOVERNMENT AND THE PRIVATE SECTOR

Because AI is a diverse field, the potential for interference with human rights depends both on the type of data a system uses and the context in which the system is implemented. For example, there are fewer and different human rights risks posed by a city government's use of AI to optimize water usage than a police department's use of a criminal risk assessment tool. With this in mind, we recommend different approaches for government and the private sector.

Recommendations for government use of AI

AI systems for government often implicate value judgments that are necessarily linked to the political process in free and democratic government systems. For this reason, and because governments can directly deprive people of their liberty, this report recommends higher standards for the public sector regarding the use of AI. States bear the primary duty to promote, protect, respect, and fulfill human rights under international law, and must not engage in or support practices that violate rights, whether in designing or implementing AI systems. They are required to protect people against human rights abuses, as well as to take positive action to facilitate the enjoyment of rights.[116] The recommendations below articulate a framework for government decision making in general, not just in AI. They apply to any type of algorithmic decision making system, regardless of whether it uses AI.

1. Follow open procurement standards. When a government body seeks to acquire an AI system or components thereof, procurement should be done openly and transparently according to open procurement standards. This includes publication of the purpose of the system, goals, parameters, and other information to facilitate public understanding. Procurement should include a period for public comment, and states should reach out to potentially affected groups where relevant to ensure an opportunity to input.

2. Mandate human rights impact assessments. States must thoroughly investigate AI systems to identify human rights risks prior to development or acquisition, as well as on a regular and ongoing basis throughout the lifecycle of the system. A human rights impact assessment may be a necessary part of a larger algorithmic impact assessment process that examines broader threats, including threats posed by uses of AI to conduct surveillance or other activity that interferes with human rights. [117] Appropriate laws must exist to govern

116 See https://www.ohchr.org/EN/ProfessionalInterest/Pages/InternationalLaw.aspx for a summary of state' human rights obligations under international law.

117 The AI Now Institute has outlined a practical framework for algorithmic impact assessments by public agencies, https://ainowinstitute.org/aiareport2018.pdf. Article 35 of the EU's General Data Protection Regulation (GDPR) sets out a requirement to carry out a Data Protection Impact Assessment (DPIA); in addition, Article 25 of the GDPR requires data protection principles to be applied by design and by default from the conception phase of a product, service or service and through its lifecycle.

the uses of AI for these purposes.[118] Any assessment process should include:

- Testing and audits by independent experts
- Identifying measures to mitigate identified risks and prevent any rights violations from occurring, and measuring compliance and efficacy
- A failsafe to terminate acquisition, deployment, or any continued use if at any point an identified human rights violation is too high or unable to be mitigated
- Identification of any new legal safeguards needed to protect human rights in specific applications of AI tools
- Special determinations of bias, particularly in the criminal justice sector due to the risks to fair trial, right to liberty, and non-discrimination
- If a third party is used to develop and/or implement the system, a requirement for the third party to participate in the human rights assessment process

3. Ensure transparency and explainability. Maximum possible transparency is necessary for any AI system, including transparency regarding its purpose, how it is used, and how it works, which must continue throughout a system's life cycle. Non-disclosure agreements and other contracts with third parties under the guise of protecting intellectual property are a violation of this principle because they prevent public oversight and accountability. Specifically, adequate transparency and explainability must include:

- Regular reporting of where and how governments use and manage AI systems
- Use of open data standards in both training data and code to the fullest extent possible, while adhering to privacy standards [119]
- Enabling independent audits of systems and data
- Clear and accessible reporting of the operation of any AI system. This means providing meaningful information about how outputs are reached and what actions are taken to minimize rights-harming impacts
- Targeted notification when a government AI system makes a decision that impacts an individual's rights
- Avoidance of "black box systems," meaning avoidance of any AI system when a person cannot meaningfully understand how it works

4. Establish accountability and procedures for remedy. The use of an AI system to do a task previously done by a human does not remove standard requirements for responsibility and accountability in government decision making processes. There should always be a human in the loop, and for high-risk areas, including criminal justice, significant human oversight is necessary. Governments should set policies regarding automation of processes, with an eye to human rights impacts. Additionally, individuals must have the right to challenge the use of an AI system or appeal a decision informed or wholly made by an AI system. More specifically, accountability and remedy require:

- Proper training for operators of an AI system. Government employees who use and manage an AI system must understand how it works, the bounds of its use, and potential for harm. Proper training ensures humans remain in the loop in a meaningful way and increases the likelihood of spotting harmful outcomes.
- Establishing responsibility for the outputs of an AI system. Although states often rely on third parties to design and implement AI systems, ultimate responsibility for human rights interferences must lie with states. To protect against abuse, government entities must acquire the technical expertise necessary to thoroughly vet a given system.
- Establishing mechanisms for appealing any given use or specific determination of an AI system.

118 See, e.g., International Principles on the Application of Human Rights to Communications Surveillance, last accessed June 15 2018, available at https://necessaryandproportionate.org/.
119 See https://www.opengovpartnership.org/sites/default/files/open-gov-guide_summary_all-topics.pdf for more information on open data standards for government data

Even systems procured and implemented transparently and with stakeholder input may still the risk of significant human rights violations.[120] To address this, there should be a process that allows the public to contest the use of an AI system in its entirety.

Recommendations for Private-Sector and Non-State Use of AI

Private-sector actors also have a responsibility to respect human rights, independent of state obligations. To meet their duty, private-sector actors must take ongoing steps to ensure they do not cause or contribute to human rights abuses.[121] The establishment of AI ethics policies by many of the large private-sector players is laudable, but a human rights impact assessment should be integrated into larger ethics review processes. Additionally, private-sector actors should pursue transparency and explainability measures, as well as establish procedures to ensure accountability and access to remedy. Collectively, this human rights due diligence, informed by expert stakeholders, assists companies in preventing and mitigating abuses. However, firms must also meet their obligations to redress harms directly or indirectly resulting from their operations, via rights-respecting processes developed in consultation with affected communities. We recognize that not all AI uses have equal risk of human rights harms, and that actions required to prevent and respond to human rights violations will depend on the context. Specifically, private-sector actors should:

1. Conduct human rights due diligence as per the UN Guiding Principles on Business and Human Rights, and consisting of the following three core steps:[122]

1. Identify potentially adverse outcomes for human rights. Private-sector actors should assess risks an AI system may cause or contribute to human rights violations. In doing this, actors must:

- Identify both direct and indirect harm as well as emotional, social, environmental, or other non-financial harm.
- Consult with relevant stakeholders in an inclusive manner, particularly any affected groups, human rights organizations, and independent human rights and AI experts.
- If the system is intended for use by a government entity, both the public and private actors should conduct an assessment.

2. Take effective action to prevent and mitigate the harms, as well as track the responses. After identifying human rights risks, private-sector actors must mitigate risks and track them over time. This requires private-sector actors to:

- Correct the system, including where risks sit with training data, design of the model, or the impact of the system.
- Ensure diversity and inclusion of relevant expertise to prevent bias by design and inadvertent harms.
- Submit AI systems with significant risk of human rights abuses to independent third-party audits.
- Halt deployment of any AI system in a context where the risk of human rights violations is too high or impossible to mitigate.
- Track steps taken to mitigate human rights harms and evaluate their efficacy. This includes regular quality assurance checks and auditing throughout the system's life cycle. This is particularly important given the role of negative feedback loops that can exacerbate harmful outcomes.

120 The state of Pennsylvania has worked hard to implement algorithmic transparency in use of automated decision making systems. However, public comments process and open data standards have not stopped problematic systems from being used. See https://slate.com/technology/2018/07/pennsylvania-commission-on-sentencing-is-trying-to-make-its-algorithm-transparent.html.

121 See UN Guiding Principles on Business and Human Rights

122 Adapted from the Toronto Declaration

3. Be transparent about efforts to identify, prevent, and mitigate the harms in AI systems. Transparency to all individuals and groups impacted as well as other relevant stakeholders is a key part of human rights due diligence, and involves communication.[123] In practice, this means that private-sector actors must:

- Publicly disclose information on identified human rights risks, including both how the system is designed or the context in which it is used.
- Publish technical details about the AI system, including samples of training data and details about sources of the data.

2. Provide transparency and explainability to the extent possible. Private-sector actors must be highly transparent and provide meaningful information about how AI systems work. Transparency is especially important when AI systems may have a significant public or personal impact, for example in medicine or in content recommendation and moderation. Specifically, this includes:

- Adherence to open source and open data standards.
- Publication of meaningful, accessible explanations of how an AI system works so that people can be meaningfully informed about how it may impact them.

3. Establish appropriate mechanisms for accountability and remedy. Private-sector actors should establish internal accountability mechanisms for the functioning of AI systems. Although states have the primary duty to provide access to formal remedy in the case of human rights violations, companies must take additional action to ensure access to meaningful, effective non-judicial remedy and redress.[124] At minimum, this includes:

- Internal responsibility for development and implementation of an AI system.
- Commitment by third parties developing AI systems for third parties to clearly delineating responsibility and accountability between vendor and client, including the vendor's obligation to ensure proper training of the risks of the system as well as to mitigate the risk of function creep and misuse of an AI system.
- Creation of clear, transparent processes by which an individual can directly submit complaints and seek redress for human rights harms in a timely manner. These could be administered internally, in collaboration with other relevant stakeholders, or through a mutually acceptable external body.[125] Findings should feed back into product and policy development to better prevent and mitigate harms.

THE NEED FOR MORE RESEARCH OF FUTURE USES OF AI

While data protection, transparency, and accountability mechanisms go far toward mitigating human rights abuses in the use of AI, they don't solve all of the foreseeable problems. For instance, as we identified, in the future, AI systems may substantially impact economic opportunities or facilitate war or conflict globally. For these reasons, we recommend that states and private-sector entities, including civil society organizations and individuals from academia, work together to investigate future uses of AI and continue to explore the potential human rights impacts identified here. Emphasis should be placed on identifying and building response mechanisms for potential threats to ensure that negative implications are mitigated to the fullest extent possible. These fora should be multi-stakeholder and pluralistic in order to ensure that all potential threats are identified and solutions don't preference any specific group over another or further diminish marginalized voices.

123 UN Guiding Principles on Business and Human Rights, Principle 21.
124 For more on the procedural and substantive aspects of remedy in the information and communications technology sector, see the Access Now "Telco Remedy Plan" at https://www.accessnow.org/cms/assets/uploads/archive/docs/Telco_Remedy_Plan.pdf.
125 See Principles 29 and 31 of the UN Guiding Principles on Business and Human Rights for more information.

REBUTTAL: TRANSPARENCY AND EXPLAINABILITY WILL NOT KILL AI INNOVATION

There are two frequently heard arguments against requirements for transparency and explainability of AI systems.[126] One argument is that AI is too complex to require transparency and doing so could damage innovation. The second argument is that explainability is impossible and, if mandated, would require AI developers to sacrifice the complexity of their systems and (again) hamper innovation. These arguments are overblown and inconsistent with developments in AI today. Each argument is addressed below.

Publishing the code and the data allows third-party experts to identify potential issues.

Opponents of transparency in AI systems question the utility of publishing the code and training data for such complex systems that may rely on millions of data points and have models that change over time.

Although auditing AI systems may present a new set of technical challenges, it does not render transparency meaningless. AI developers vet training data and AI outputs to identify sources of bias and test for equitable outcomes. Transparency would ensure access to necessary training data and AI outputs for independent experts to identify sources of bias and test for equitable outcomes, including identifying any problems developers missed or masked. This process necessarily increases accountability and fosters user trust.[127]

Full transparency is vital in high-risk cases and need not damage companies.

Opponents of transparency also argue that transparency requirements reduce incentives to invest in new AI systems because it would allow replication.

While history has shown that open source projects are often not only successful but facilitate innovation, there may be an option in limited circumstances where private-sector actors determine that route is untenable. In such cases, private-sector actors could facilitate access to relevant code by trusted, identified third parties for audits and testing. Recently, Facebook has given select researchers access to data to study election interference on the platform.[128] Given increased public scrutiny of the role of algorithms in our lives, it is conceivable that other companies will follow suit. However, information on data sets used and outputs should still be published, as well as any other information that could facilitate understanding and measure bias.

However, states should always be required to provide full transparency for government use of AI systems. This is particularly important in areas such as in law enforcement and the justice system. Although companies supplying the software may have reasons for not publishing the code and the training data in these cases, fundamental human rights cannot be sacrificed for the sake of corporate interests.

Meaningful explainability is increasingly possible, and the AI world is largely behind it.

Opponents of regulation argue that requiring explainability would mean systems would have to be substantially less complex, and therefore less accurate, ultimately stifling innovation. This argument misses the mark in a number of ways.

First, valuable levels of explainability are achievable. Although Facebook may not fully understand how its targeting advertising algorithm works, it knows enough to tell users what actions led to them being served a certain advertisement. This type of information matters, and it is easy to provide. Research suggests that it is

126 See Joshua New, "How (and how not) fo fix AI," Tech Crunch, July 26, 2018, https://techcrunch.com/2018/07/26/how-and-how-not-to-fix-ai/?mc_cid=b4840b1adb&mc_eid=[UNIQID].
127 See, e.g., https://motherboard.vice.com/en_us/article/a34pp4/john-deere-tractor-hacking-big-data-surveillance ("The American Farm Bureau helped construct the "Privacy and Security Principles for Farm Data," which addresses issues of data ownership, portability, use, and sharing. Companies like Deere and Monsanto were early signers, but questions remain about how much these principes protect in practice.").
128 https://www.accessnow.org/cms/assets/uploads/2018/07/GDPR-User-Guide_digital.pdf

even possible for systems to measure how a given input affected the output. In a system used by universities to rank applicants, this could tell you, for example, that 20% of the ranking is due to GPA, 25% due to standardized testing, and 10% due to school ranking and other factors. This level of explanation would go a long way in addressing the "black box" challenge and identifying potential sources of bias.

Additionally, explainability is technically valuable. Developers need to be able to determine whether a system is solving the right problem. There are many examples of AI systems that "cheated" to arrive at the desired outcome. For example, researchers at the University of Washington created a deliberately bad algorithm that was supposed to classify images of husky dogs and wolves. The system correctly labeled the images, but rather than learning the difference between the appearance of huskies vs. wolves, the system detected the presence of snow because most of the images of wolves had snow in the background.[129] If AI systems in high stakes fields ultimately solve the wrong problem, the outcome could be life threatening.

Because explainability is necessary for the adoption of AI in certain fields, in some ways the quest for explainability is spurring AI innovation. For both ethical and technical reasons, academics and major AI companies alike are devoting significant effort toward explainability, and they are making serious progress. In August 2018, Google's DeepMind published a study about an AI system it developed to identify eye disease in 3D ocular scans. When the system makes a diagnosis, it points out the portions of the scan it used so that physicians can see how it arrived at that diagnosis, as well as how confident it is in the diagnosis.[130] Breakthroughs such as this show how explainability may be driving innovation in AI.

VIII. CONCLUSION

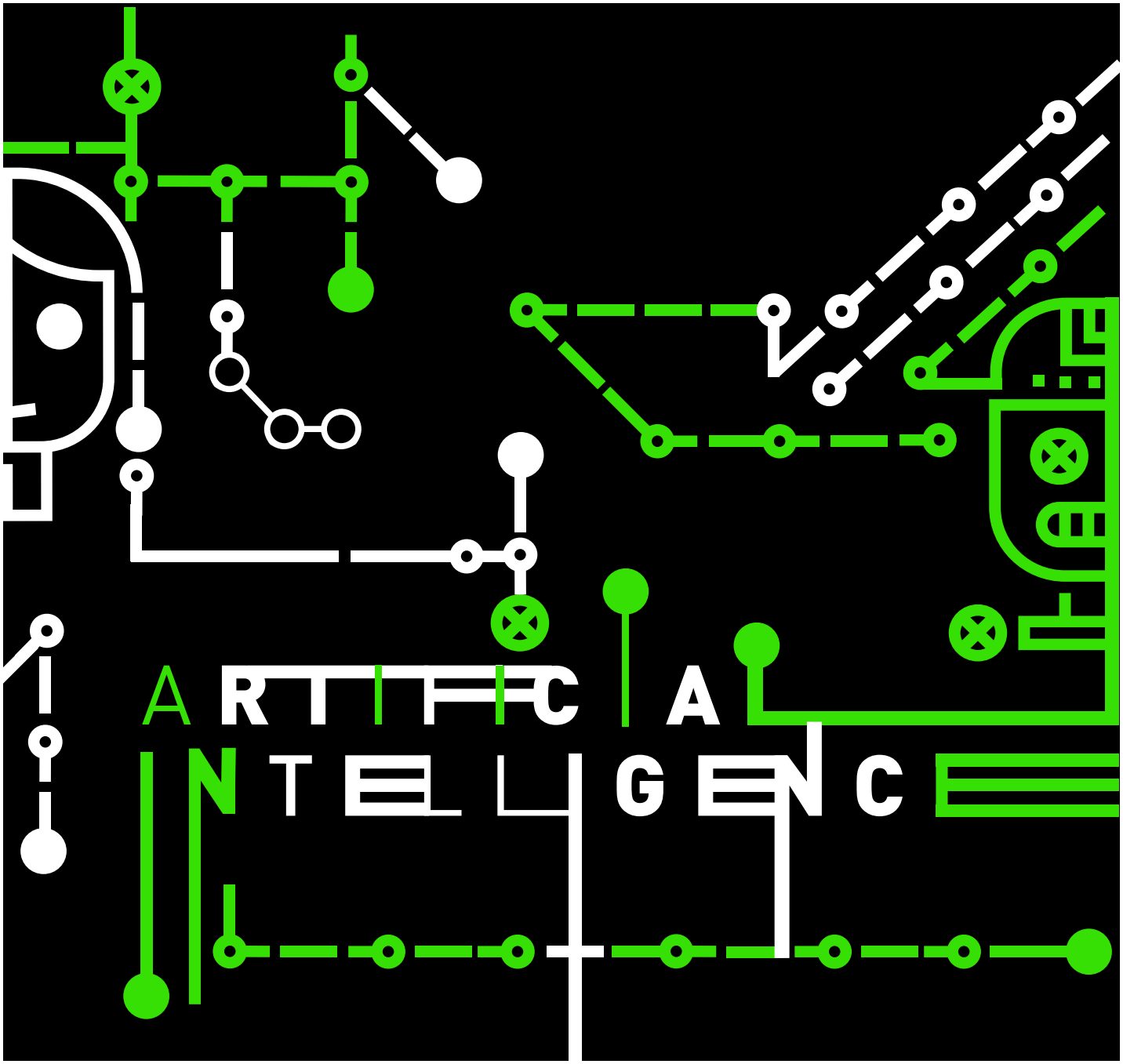
Artificial intelligence systems are changing the way things are done in companies and governments around the world, and bringing with them potential for significant interference with human rights. Data protection laws and safeguards for accountability and transparency, like those we have described in this paper, may be able to mitigate some of the worst uses known today, but more work is necessary to safeguard human rights as AI technology gets more sophisticated and expands into other areas. We hope this report helps to inspire deeper conversations in this crucial area for those who care about the future of human rights, and we look forward to engaging in those conversations.

129 In the U.S., many of the details of government use of algorithmic decision making system are hidden behind non-disclosure agreements and memorandums of understanding with vendors. See https://www.wired.com/story/when-government-rules-by-software-citizens-are-left-in-the-dark/.
130 See https://www.ohchr.org/EN/ProfessionalInterest/Pages/InternationalLaw.aspx for a summary of state's human rights obligations under international law.

Access Now defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age.

For more information, visit **https://www.accessnow.org** or contact **info@accessnow.org.**

Nov, 2018



HUMAN RIGHTS IN THE AGE OF ARTIFICIAL INTELLIGENCE

accessnow