ARTIFICIAL INTELLIGENCE: THE CONSEQUENCES FOR HUMAN RIGHTS

HEARING

BEFORE THE

TOM LANTOS HUMAN RIGHTS COMMISSION

HOUSE OF REPRESENTATIVES

ONE HUNDRED AND FIFTEENTH CONGRESS

SECOND SESSION

MAY 22, 2018

Available via the World Wide Web: humanrightscommission.house.gov

TOM LANTOS HUMAN RIGHTS COMMISSION

RANDY HULTGREN, Illinois,
Co-chairman

JAMES P. McGOVERN, Massachusetts,
Co-chairman

GUS BILIRAKIS, Florida
BARBARA COMSTOCK, Virginia
DARIN LaHOOD, Illinois

KEITH ELLISON, Minnesota
TED LIEU, California
JAN SCHAKOWSKY, Illinois
NORMA J. TORRES, California

JAMIE STALEY, *Senior Professional Staff*
MATTHEW SINGER, *State Department – Capitol Hill Fellow*

KIMBERLY STANTON, *Senior Professional Staff*
ROSIE BERMAN, *Democratic Intern*

C O N T E N T S

WITNESSES

LETTERS, STATEMENTS, ETC., SUBMITTED FOR THE HEARING

APPENDIX

ARTIFICIAL INTELLIGENCE: THE CONSEQUENCES FOR HUMAN RIGHTS

TUESDAY, MAY 22, 2018

HOUSE OF REPRESENTATIVES,
TOM LANTOS HUMAN RIGHTS COMMISSION,
Washington, D.C.

The commission met, pursuant to call, at 3:00 p.m., in Room 2255 Rayburn House Office Building, Hon. James P. McGovern and Hon. Randy Hultgren [co-chairmen of the commission] presiding.

Mr. HULTGREN: We are going to go ahead and get started. I think some of my colleagues will be coming in. It's a busy day with markups and rules and some other things. So my sense is people might be coming in and out a little bit.

But I know you all have busy schedules as well so and they will be calling votes for us probably in a little over an hour.

Good afternoon. I want to welcome everybody to the Tom Lantos Human Rights Commission's hearing on artificial intelligence and the consequences for human rights. As a reminder to those in the audience, I want to encourage you to please turn off your phones and electronic devices or set them to silent.

My interest in today's topic is two-fold. As co-chair of the commission, I am interested in key trends in human rights, and as a member of the House Subcommittee on Research and Technology, I am also a watcher of emerging technologies such as artificial intelligence.

Much has been written about artificial intelligence, or AI for short. AI is not one single technology but a whole new class of programs that will fundamentally change how computers process information.

Even though AI technology is relatively new, it's already profoundly affecting fields as diverse as health care, education, law enforcement, sales, and many others.

In the right hands, AI technologies have the power to do profound good by saving lives and reducing the cost of essential services.

As some of our panelists will note, AI even has potential to be a powerful tool to help advance the work that human rights defenders are doing.

Unfortunately, as we have learned time and again, there is no such thing as a technology that is exclusively used for good causes. In the wrong hands, AI has the potential to negatively affect many aspects of our lives, and that does include human rights.

In remarks I recently delivered to the European Parliament's subcommittee on Human Rights, I stated, and I quote, "There have been numerous press reports of the massive deployment of surveillance technologies against Uighurs in China, including the use of artificial intelligence software and facial recognition software," end quote.

The Chinese government's tactics against the Uighur people in Xinjiang Province have been a laboratory for cutting-edge surveillance technology that truly is Orwellian and the repercussions of that misuse of technology could extend far beyond China.

This hearing is designed to discuss those broader repercussions and to explore ways that these threats can be, if not controlled or totally prevented, at least countered in meaningful ways because while many have already begun to discuss ways to prevent the unethical use of AI from a national security standpoint, these conversations have rarely addressed international human rights.

Recent news items about the misuse of social media and other internet tools to gather massive amounts of information and make surprisingly accurate predictions based on that information have raised privacy concerns.

Some of that misuse involves AI tools. These same tools could be used by abusive regimes to single out political opponents or to track and harass human rights defenders.

And once the AI genie is out of the bottle, it could give nonstate actors an unprecedented ability to commit human rights violations normally associated with national governments.

That raises some of the fundamental questions of this hearing. What are the ways that AI could be abused to violate internationally recognized human rights?

Are there ways to prevent AI being used for such abuses? What is the role of the U.S. government or any government to either prevent or mitigate the use of AI for repressing human rights?

What are the responsibilities and what's the role of the technology industry to prevent the misuse of their technologies for human rights abuses?

Even if governments and the tech industry try to prevent or counteract human rights abuses enabled by artificial intelligence, does the borderless nature of the modern internet doom such efforts to only a limited scope?

I would like to thank our distinguished witnesses for coming today to address these concerns. We do appreciate your presence here as we discuss these important topics.

Amnesty International could not send a panelist today, but I thank them for submitting a statement for the record that, among other things, describes the Toronto Declaration, which addresses the risk of human rights harms associated with AI.

I would also like to thank the Future of Humanity Institute at Oxford University in England, which has submitted a statement for the record. These statements will be available on the commission's website.

Mr. HULTGREN: With that, I'd like to move to our panelists and I am grateful, again, for their willingness to be with us today.

First, we have got Samir Goswami, who is -- works with NGOs, government alliances, and businesses to develop technology, enabled ethics supply initiatives and programs. He's an independent consultant with past and current clients that include the United Way Worldwide, Organization for Security and Cooperation in Europe, the Ethical Trading Initiative, City of Houston, Humanity United, ISRA Institute, and Mobile Accord. Thank you for being here.

We also have Paul Scharre here, a senior fellow and director of the technology and national security program at the Center for a New American Security. He is the author of "Army of None: Autonomous Weapons in the Future of War," which was published just very recently in April of 2018.

Also grateful to have Professor Kenneth Anderson, who teaches and writes in the area of business and international law, public international law and governance, law of war and armed conflict and, most recently, law and regulation of emerging technologies, particularly automation, robotics, and AI.

He's published extensively on national security law topics, particularly counterterrorism, drone warfare, and autonomous weapons, and he also serves as book review editor of the National Security and Law website, Lawfare. So with that -- and he is a professor of law at American University.

So, again, thank you to each of our panelists. I am going to ask each of you if you would present your testimony and then we will move to questions at that time.
So Mr. Goswami, if you would start us. Can you make sure your microphone is on or pull it good and close?

[The prepared statement of Co-chair Hultgren follows]

PREARED STATEMENT OF THE HONORABLE RANDY HULTGREN, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF ILLINOIS AND CO-CHAIRMAN OF THE TOM LANTOS HUMAN RIGHTS COMMISSION

Tom Lantos Human Rights Commission Hearing

Artificial Intelligence: The Consequences for Human Rights

May 22, 2018
3:00 – 4:30 PM
2255 Rayburn House Office Building

Opening Remarks as prepared for delivery

Good afternoon, and welcome to the Tom Lantos Human Rights Commission's hearing on Artificial Intelligence and the consequences for Human Rights. My interest in today's topic is two-fold: as Co-Chair of this Commission, I am interested in key trends in human rights, and as a member of the House Subcommittee on Research and Technology, I am also a watcher of emerging technologies such as artificial intelligence.

Much has been written about artificial intelligence, or "AI" for short. AI is not one single technology, but a whole new class of programs that will fundamentally change how computers process information. Even though AI technology is relatively new, it already profoundly affects fields as diverse as health care, education, law enforcement, sales and many others.

In the right hands, AI technologies have the power to do profound good by saving lives and reducing the cost of essential services. As some of our panelists will note, AI even has potential to be a powerful tool to help advance the work that human rights defenders are doing.

Unfortunately, as we have learned time and again, there is no such thing as a technology that is exclusively used for good causes: in the wrong hands, AI has the potential to negatively affect many aspects of our lives, and that includes human rights.

In remarks I recently delivered to the European Parliament's subcommittee on Human Rights, I stated "There have been numerous press reports of the massive deployment of surveillance technology against Uyghurs in China, including the use of Artificial Intelligence software and facial recognition software." The Chinese government's tactics against the Uyghur people in Xinjiang Province have been a laboratory for cutting edge surveillance technology that is Orwellian – and the repercussions of that misuse of technology could extend far beyond China.

This hearing is designed to discuss those broader repercussions, and to explore ways that these threats can be – if not controlled or totally prevented – at least countered in meaningful ways. Because while many have already begun to discuss ways to prevent the unethical use of AI from a national security standpoint, these conversations have rarely addressed international human rights.

Recent news items about the misuse of social media and other internet tools to gather massive amounts of information – and make surprisingly accurate predictions based on that information – have raised privacy concerns. Some of that misuse involved AI tools. These same tools could be used by abusive regimes to single out political opponents. Or to track and harass human rights defenders. And once the AI "genie" is out of the bottle, it could give non-state actors an unprecedented ability to commit human rights violations normally associated with national governments.

That raises some of the fundamental questions of this hearing:

- What are the ways that AI could be abused to violate internationally recognized human rights?
- Are there ways to prevent AI being utilized for such abuses?
- What is the role of the U.S. government – or any government – to either prevent or mitigate the use of AI for repressing human rights?
- What are the responsibilities and role of the tech industry to prevent the misuse of their technologies for human rights abuses?

Even if governments and the tech industry try to prevent, or counteract, human rights abuses enabled by artificial intelligence, does the borderless nature of the modern internet doom such efforts to only a limited scope?

I would like to thank our distinguished witnesses for coming today to address these concerns. We appreciate your presence here as we discuss this important topic.

Amnesty International could not send a panelist today, but I thank them for submitting a statement for the record that, among other things, describes the "Toronto Declaration," which addresses the risk of human rights harms associated with AI. I would like to thank the Future of Humanity Institute at Oxford University in England, which has also submitted a statement for the record. These statements will be available on the Commission's website.

STATEMENTS OF SAMIR GOSWAMI, CONSULTANT, 3RD PARTY LLC; PAUL SCHARRE, SENIOR FELLOW AND DIRECTOR, TECHNOLOGY AND NATIONAL SECURITY PROGRAM, CENTER FOR A NEW AMERICAN SECURITY; KENNETH ANDERSON, PROFESSOR, AMERICAN UNIVERSITY

STATEMENT OF SAMIR GOSWAMI, CONSULTANT, 3RD PARTY LLC

Mr. GOSWAMI: Thank you, Chairman Hultgren, and thank you for inviting me to testify today.

While AI has great potential to uphold and promote human rights, conversely it can also be used to suppress it. The primary thought for consideration is that while AI has tremendously improved our ability to process the world around us, we don't often act upon the insights that we glean from it.

That is, while machines may help us understand problems and human rights issues better, we, the humans, have develop the political will to intervene.

Unfortunately, this is something we don't do enough of in the human rights space. As a collection of technologies that involve the processing of very large amounts of data to machine learning, which is a core component of AI, machines can be programmed to imitate certain ways our human brains process information.

That is, the machine can be taught to observe, identify, and classify. It could even be taught to make mistakes and learn from those mistakes.

AI can be used to uncover human rights violations that many workers face around the world in the factories, farms, and mines that they labor in.

The U.S. Department of Labor finds that 139 goods from 75 countries may be made from childhood forced labor. A U.S. company may have thousands of suppliers around the globe and all vary on how they treat their workers.

Most U.S. companies have codes of conduct that they expect each one of those suppliers to abide by and use audits to verify that those factories are indeed doing so. These audits, close to 50 to 100 pages per factory, multiplied by thousands of factories around the world, generates lots of data.

Supply chain managers can use computing technology to process this vast amount of audit data to flag issues. However, these audits can be forged or they could be susceptible to other influences.

The erroneous data can contaminate an AI-enabled analysis that may not then paint a complete and accurate picture of whether the supplier is acting ethically.

AI-enabled systems can conduct outside validation to complement audit data through accessing and processing other information sources such as news reports, court filings, public records, and other materials associated with that supplier.

Furthermore, workers can also leave a data footprint and AI can be used in real time to scan social media chat rooms, message boards, or public comment websites for any references to those suppliers left by the workers.

All these various streams of data can be analyzed together to provide an independent human rights assessment of the supplier's labor practices.

However, this capability in this example can also be flipped for illicit purposes. Machine learning and AI can be used to comb through workers' social media posts to target union organizers or those the state or factory owner may deem to be a trouble maker.

Facial recognition technology can be coupled with AI to find and target migrant workers or human rights defenders who are challenging repressive labor regimes, and predictive capabilities might flag workers and subject them to arbitrary detention or harassment based on the AI-informed suspicion that they might challenge employment practices or poor working conditions.

In addition, we also need to ensure that AI does not just generate wealth for a few at the expense of others. Increased automation can lead to a decrease in certain types of jobs, displace low-wage workers, and depress wages.

A 2016 White House report finds that anywhere from 9 percent 47 percent of jobs over the next two decades could be disrupted by AI and automation. These impacts will be borne on the shoulders of low-income women and migrants, who are already some of the world's most vulnerable.

The federal government should invest in AI for good. However, we also need to act upon the insights that AI and machine learning deliver to us. Our investment criteria should not just be that the technology was developed or deployed.

We should measure the human outcomes that were achieved -- that is, to the good that we were seeking in an AI for good application actually happen.

For example, we can use AI to help pinpoint exactly which factory might utilize child labor. But that insight is wasted if we don't respond and we deliberately ignore our ethical obligations to those children.

Unfortunately, while some progress is being made, U.S. companies are simply not doing enough to act upon the technology-enabled insights on labor abuses that they have access to.

There is, thus, tremendous opportunity for the U.S. Department of Labor, the USAID, the State Department, DOJ, CPB, and other entities to use AI and machine learning to verify how workers tied to U.S. public and private supply chains are being treated.

These insights can be used to apply laws already on the books that prohibit forced and child labor-made goods to enter the U.S. or to enforce trade agreements that have often ignored labor practices in place.

The U.S. government needs to act upon such technology-gleaned insights by compelling companies to drive supply chain improvements, enable law enforcement to prosecute those who abuse human rights, and press other governments to uphold workers' rights.

Civil society and industry are also taking steps to address these issues. For example, the International Corporate Accountability Roundtable is mapping the disruptions to labor markets that automation and robotics can lead to and the partnership on AI is bringing the technology industry together with civil society and human rights organizations to collectively identify solutions and safeguards to various AI influence challenges.

In conclusion, I strongly believe that AI has and can have a tremendous impact on human rights. We need to ensure that the wealth that AI will generate will be shared broadly and not exacerbate existing economic disparities.

The application of AI should also be measured by the outcomes it produces and whether those violate human rights principles.

Most of all, we need to act upon the insights we glean. Technology is just a tool to help us understand a problem better, which is not a replacement for the political will that is needed to drive change.

Thank you for your time and leadership and the opportunity to address this commission.

[The prepared statement of Samir Goswami follows]

PREPARED STATEMENT OF SAMIR GOSWAMI

3rd Party LLC

TESTIMONY OF
Samir Goswami, Consultant, 3rd Party LLC

HEARING ON:
Artificial Intelligence: The Consequences for Human Rights
Before the Tom Lantos Human Rights Commission

May 22, 2018

Chairman Hultgren, Chairman McGovern and Members of the Commission, thank you for inviting me to testify today to discuss the implications of Artificial Intelligence, or AI, on human rights. AI is proving to be tremendously beneficial in the transportation, logistics, health care, defense, military intelligence and other sectors; and, while AI is showing great potential to uphold and promote human rights, conversely, it can also be used to suppress it. Today, I will provide examples and conclude with some initiatives already underway to help us better understand and guide AIs implications. However, an important point I want to leave you with is that while AI has tremendously improved our ability to process the world around us, we don't often act upon the insights we glean. That is, while machines may help us understand problems and human rights issues better—we, the humans, have to develop the political will to intervene; unfortunately, this is something we don't do enough of in the human rights space.

A. Introduction & Background:

Conceptually, AI is about building machines that are capable of thinking like, or at least mimicking the thinking processes of humans. Forbes Magazine states, "*AI can be thought of as simulating capacity for abstract, creative, deductive thought, and particularly the ability to learn*." If you have ever asked Siri a question, applied for a credit card online, or ordered a product through Alexa, you have most likely interacted with AI. In practical terms, AI is a collection of technologies that involve the processing of very large amounts of data. For example, banks can use AI to detect fraud patterns by analyzing millions of financial transactions; large grocery stores may use AI to accurately predict customer buying preferences resulting in better control of inventory and less waste. During flight, an airplane's engines may send a constant stream of performance data to a central server, which will analyze the information with other data, such as its age, routes it has flown, weather conditions it has encountered, and even records of the experiences of the pilot. It can process this information and provide an analysis to a human engineer containing a snapshot of system health, flagging anomalies or predicting which parts may need servicing.

Through a process referred to as "machine learning", a core component of AI, machines can be programmed to imitate certain ways our human brains process information; that is a machine can be taught to observe, identify and classify. It can even be taught to make and learn from its mistakes. This incredible progress is possible because of the massive amounts of data that we generate and the improvements in computing power that enables machines to quickly process all the information at its disposal—greatly expanding our ability to solve problems and understand the world around us.

B. Opportunities, Responsibilities and Applications for Labor and Human Rights

Amnesty International is piloting the use of machine learning and AI in their human rights investigations and response. When I was a managing director at Amnesty International USA we had access to over 30 years of meticulously recorded human rights data. Our goal was to test if we could use this historic data coupled with records of current occurrences to predict which of the 500 human rights incidents that we tracked every year needed the most attention from Amnesty's campaigners. Through a partnership with Purdue University and the nonprofit DataKind, we had volunteer computer scientists and programmers develop algorithms that sorted through 1.4 million lines and 11,000 files of data—all in a matter of hours. They were able to create a preliminary model that at least in our tests, correctly predicted a binary outcome with over eighty percent accuracy. This type of proactive analysis could enable human rights organizations to create heat-maps of urgency, warn human rights defenders of the severity of risks, or alert first responders to deploy interventions. Amnesty International continues groundbreaking research into AI applications for human rights and has also submitted a statement for the record to this Commission.

AI is also being used effectively to help counter human trafficking. The California based nonprofit organization Thorn is helping law enforcement identify human traffickers through machine learning and AI applications. Traffickers will often advertise the availability of their victims through classified advertisements, for example for escort or other adult services. The same trafficker or trafficking entity may use multiple phone numbers across hundreds of

different advertisements. To find connections and link victims to the same trafficker, law enforcement would have to manually scroll through thousands of these ads that are updated daily, which simply is too much data for a human to process. Thorn is using machine learning and AI to scan the web and the dark web to recognize common identifiers such as phone numbers, or similar styles of writing in such advertisements, and linking that data to the digital footprint that traffickers may leave.

AI is also being used to uncover human rights violations that many workers face around the world in the factories, farms and mines that they labor in. The U.S. Department of Labor finds that 139 goods from 75 countries may be made from child or forced labor. A U.S. company may have thousands of suppliers around the globe that provide raw materials, labor and other services to produce the goods that we consume. Most U.S. companies have codes of conduct that establish labor standards that they expect each one of their suppliers to abide by. However, each one of these suppliers operates in its own legal and social environments and has varying labor practices. Many U.S. companies will use on the ground audits to verify that the supplier is indeed complying with its standards. These audits generate a lot of data—close to 50 to 100 pages per factory, multiplied by thousands of factories, at least once a year.

Supply chain managers can use machine learning and AI to process this vast amount of audit data to flag issues. However, audits can be forged or be susceptible to other influences—this erroneous data can contaminate an AI enabled analysis that may then not paint a complete and accurate picture of whether the supplier is acting ethically and workers' rights are being upheld. AI systems can be used to conduct outside validation to complement audit data through accessing and processing other information sources, such as news reports, court filings, public records, any materials that compromise the open source data footprint of a supplier and its business associates. Furthermore, workers also leave a data footprint—machines can scan social media, chat forums, message boards or public comment websites for any references about those suppliers made by workers. Mobile phones can be used to deploy surveys to workers directly, independent of supplier supervision, data from which can also be

incorporated into an analysis. All these various streams of data can and are being analyzed together by an AI enabled system to provide an independent human rights assessment of a supplier's labor practices.

It is important to note that all of this data and computing power are only useful if we actually act upon the AI informed insights we glean—which we don't often do, a point I will expand upon in the recommendations section of this testimony.

C. Consequences

I submit three broad categories of consequences for the Commission to consider. This is by no means an exhaustive list:

1. First, while AI can be a powerful tool in the labor rights space, the above example also leads to the obvious risks that AI can pose; that is, this level of analysis can also be flipped for illicit purposes. For example, machine learning and AI can be used to comb through worker social media posts to identify union organizers or those a State or factory owner may deem to be a "trouble maker". Facial recognition technology can be coupled with AI and machine learning to target migrant workers or human rights defenders who are challenging repressive labor regimes; and, predictive capabilities might flag workers and subject them to arbitrary detention or harassment based on the AI informed suspicion that they *might* challenge employment practices and poor working conditions in the future. While I may be able to point to numerous pilot projects that utilize AI to specifically advance human and labor rights—rogue states and actors can use technology enhanced by machine learning and AI to suppress civil and human rights at considerable scale.

2. Second, those developing AI systems have to be very aware of the human prejudices that the machine may inherit. AI is only as good as the data it learns from—if the data has biases they will be amplified by the machines. For example, if a computer is ingesting large quantities of employment data to inform an algorithm that selects candidates for a high

paying job, and that data, because of historic biases, mainly contains data sets of men, this may lead to programmed gender biases that can exacerbate discrimination against women.

3. Lastly, and importantly, AI can be used to generate wealth for a select few at the expense of others. Increased automation can lead to the decrease in employment, displace low wage workers and depress wages. A recent publication by the Council on Foreign Relations states, "*Accelerating technological change, including automation and advances in artificial intelligence that can perform complex cogitative tasks, will alter or replace many human jobs.*" A 2016 White House report on AI and the economy states that "*Because AI is not a single technology, but rather a collection of technologies that are applied to specific tasks, the effects of AI will be felt unevenly through the economy*." This White House report finds that anywhere from nine percent to forty seven percent of jobs over the next two decades could be disrupted by AI and automation. This also holds true for U.S. companies with overseas supply chains. For example, the International Corporate Accountability Roundtable estimates that two-thirds of all jobs in developing countries could face significant automation, mainly in the apparel, electronics and agricultural sectors. A 2016 report by the International Labour Organization identifies the risks that "automation, robots and artificial intelligence" will place on millions of workers in Asia. These impacts will be born on the shoulders of low-income women and migrants who are already some of the world's most vulnerable.

   However, these impacts have not been realized just yet, and can still be mitigated. The Council on Foreign Relations also states that, "*In the absence of mitigating policies, automation and artificial intelligence are likely to exacerbate inequality and leave more Americans behind.*" We thus need to ensure that as technology is helping us make improvements in the marketplace to how we produce, distribute and consume things—we are not further exacerbating existing discriminatory tendencies, or making things worse for those already vulnerable—all grave human rights concerns.

D. Considerations & Recommendations:

Given these consequences, I submit the following considerations and recommendations:

1. First, the federal government should invest in "AI for good" and provide seed funding for such applications. However, we also need to act upon the insights that AI and machine learning deliver to us. Our investment criteria shouldn't just be that the technology was developed or deployed—we should measure the human outcomes that actually happened. That is, did the "good" that we envisioned in an "AI for good" technology application actually happen? A machine can help us better understand, but humans have to intervene. AI may help us find and understand a problem better and get down to the most accurate data point—but we still have to act upon it. For example, we can use AI to help pinpoint exactly which factory might utilize child labor—but that insight is wasted if we don't respond, and we deliberately ignore our ethical obligations to those children. Unfortunately, while some progress is being made, U.S. companies are simply not doing enough to act upon the technology enabled insights on labor and human rights abuses that they have access to.

   Additional policies and public and political pressure are needed to compel companies to actively monitor their supply chains for human rights abuses, and increased legal accountability is needed for those who don't. There is thus tremendous opportunity for the U.S. Department of Labor, USAID, DOJ, CBP and others to use AI and machine learning to verify how workers tied to U.S. public and private supply chains are being treated. These insights can be used to apply laws already on the books that prohibit forced labor and child labor made goods to enter the U.S., or to enforce trade agreements that have often-ignored labor protections in place. The U.S. government needs to act upon such technology-gleaned insights by compelling companies to drive supply chain improvements, enable law enforcement to prosecute those who abuse human rights, and press other governments to uphold workers' rights. In the supply chain management field, even without AI there is ample technology already available to determine if a supplier is treating its workers fairly; however, both governments and companies don't often act upon it.

2. Second, as the 2016 White House report on AI stated, "*Whether AI leads to unemployment and increases in inequality over the long-run depends not only on the technology itself but also on the institutions and policies that are in place*." Thus, at a macro level we need to ensure that we are properly preparing and training our workforce to avail of new technology jobs, and meaningfully assist in transitioning those who are at risk of being displaced. This includes workers in the territorial United States and those whose livelihoods are tied to the supply chains of U.S. companies overseas, with a particular focus on women and migrant workers who often conduct low wage work. Unfortunately, our collective track record is not great on these fronts. Employment loss due to technological advancements has happened on many occasions, while U.S. investments in labor market programs, such as job readiness and high-skills training, has decreased significantly over time, and is far less than those made by other industrialized countries.

3. Third, U.S. corporations developing and deploying AI need to incorporate a rights-based approach. AI systems need to be designed in ways that don't replicate human biases. Engineering teams designing AI systems need to be diverse, and the data that feeds into their systems also have to be corrected for biases. The Global Future Council on Human Rights recommends four central principles to combat bias and uphold human rights in machine learning: active inclusion, fairness, the right to understanding and access to remedy. For example, a consumer should be informed if AI was used to influence a decision about their lives (e.g. whether you get a mortgage), and should have access to a process for redress for erroneous or biased interpretations. The Center for Data Innovation recommends the concept of "algorithmic accountability", which they define as "*the principle that an algorithmic system should employ a variety of controls to ensure the operator can verify it acts in accordance with its intentions, as well as identify and rectify harmful outcomes*". In short, businesses need to ensure that the AI systems they create and utilize are not creating value for their shareholders and customers at the expense of human and civil rights.

E. Multi-stakeholder Initiatives to Understand AI and its Consequences on Human and Civil Rights

Overall, I believe that AI and other technologies have and can have a tremendous positive impact on human rights, and we need to prepare ourselves for the resulting implications in a collaborative manner. Already we are seeing progress that should be continued and supported. For example:

1. The Council on Foreign Relations, through their "The Work Ahead" project makes numerous recommendations including the need to strengthen the link between education and work through increased investments. The Council also calls on the nation's governors, Congress and the Administration to collectively establish a process to understand and address such technology implications.

2. The International Corporate Accountability Roundtable, a Washington DC based civil society organization through its "Robots and Rights" project is mapping sectors that rely heavily on low-skill human labor and how they will be impacted by automation and mechanization with a report to be released next month. The report will include "*policy and advocacy strategies and solutions from the perspective of both States and companies*".

3. Finally, the Partnership on AI, a San Francisco based non-profit association has been founded and funded by leading companies that are at the forefront of developing and applying AI technologies. The Partnership has brought together companies including Google, Amazon, Facebook, eBay and others with nonprofit civil and human rights organizations such as the ACLU, Amnesty International, the Center for Democracy and Technology, and the Electronic Frontier Foundation, to collectively identify solutions and safeguards to various AI influenced challenges. The Partnership aims to develop analysis and recommendations on how AI influences labor, the economy, and social good among other topics.

F. Conclusion:

Like any technology, AI has both positives and negative applications and effects. With the exponential surge in the availability of data and computing power comes an increase in our reliance on machines to help us do things we couldn't imagine a few years ago. Better data collection and analysis, finding patterns of human rights violations in large data sets, enabling quicker response to human rights incidents, all increase our ability to help one another. However, we have to choose to do so and not just limit AIs potential to commercial applications whose sole purpose is to increase wealth. And, AI will generate wealth—we need to ensure that that this prosperity will be shared broadly and not exacerbate existing economic disparities.

We need to proactively guide AIs myriad applications and prepare ourselves for the resulting implications in a collaborative manner. The development and deployment of AI technologies has to be within a policy framework that takes human rights principles into account, and the application of these technologies has to be matched with policies and programs that adequately prepare those who might stand to lose the most.

Most of all, we need to act upon the insights we glean: technology is just a tool to help us understand a problem better—it is not a replacement for the political will that is needed to drive change.

Thank you for your time and leadership and the opportunity to address this commission.

Mr. HULTGREN: Mr. Goswani, thank you so much.

Next, Mr. Scharre.

STATEMENT OF PAUL SCHARRE, SENIOR FELLOW AND DIRECTOR, TECHNOLOGY AND NATIONAL SECURITY PROGRAM, CENTER FOR A NEW AMERICAN SECURITY

Mr. SCHARRE: Thank you, Chairman Hultgren, for inviting me to testify today. Recent years have seen rapid advances in artificial intelligence and machine learning. AI tools are now being applied to a range of industries and will have similar applicability to human rights.

Artificial intelligence is a general purpose enabling technology much like electricity or computers. AI tools can be used for a variety of applications including data classification, anomaly detection, prediction, and optimization.

These tools will be used by state and nonstate actors for a variety of purposes, some of which will no doubt include suppressing human rights.

Other uses may help to enhance human rights or fight against repressive regimes. It is not the intention today to estimate what the net effect of AI technology will be for human rights.

Rather, I would like to walk through some potential use cases to illustrate some of the possibilities as AI technology becomes more widely used.

In the hands of a repressive state with access to large data sets about its population, AI tools could be used to increase state control. Automated facial recognition technology, combined with security cameras could make 1984-style continuous monitoring feasible.

Combined with readily available digital data collected through computers and smart phones, AI tools could be comprehensively used to monitor a person's behavior, communications, likes, and desires at a scale that not even Orwell could have imagined. People living in the digital age create a cornucopia of data. Smart phones yield location, browser history, web search history, online purchases, contacts, social media engagement, email and text message content, telephone calls, and more.

Whoever has access to this data has tremendous insight not only into a person's past but also the ability to predict their future behavior.

Without AI tools, however, it is hopelessly impractical to manage this data at scale. Artificial intelligence makes much of this data more discoverable through data

classification tools that can recognize faces, identify human emotions, translate voice to text, translate languages and process language.

AI tools also make it feasible to analyze and process this data at scale. This means that the kind of intrusive monitoring that would have in the past been extremely time consuming and resource intensive can now be done quickly and at scale along more extensive and intrusive monitoring of a population.

Moreover, large data sets can be aggregated to generate statistically valid predictions. By learning from data across an entire population and then applying this to readily available information about an individual, AI tools could be used to make predictions about that individual's preferences or behavior -- political, financial, sexual, or other.

AI tools could be used to not only monitor a population but predictably crack down on would-be dissidents.

At the same time, there are a number of features of AI tools that would make them very powerful allies for those fighting repressive regimes.

AI systems embed expertise within software itself, lowering the bar for the skills needed for a given capability.

One does not need to spend years learning chess anymore to play at the level of a grand master, for example. One can merely download a chess app for free.

Similarly, AI systems will put greater abilities in the hands of nonstate groups and individuals. Smart phones already turn surveillance tools against the state, allowing citizens to record abuses by authorities.

AI tools such as embedded object recognition or facial recognition in the hands of everyday citizens can make it even easier to identify abusers and hold perpetrators to account.

A core feature of information technology is that it renders the cost of copying and transmitting information close to zero. One of the consequences of this is that it is difficult to keep information secret.

While this is true for personal information, it is also true for state secrets. Individuals have accessed and released large tranches of government secrets on a scale that was impossible in a predigital era.

The ease with which information freely flows in the digital age is a hindrance to repressive regimes that thrive on secrecy.

AI tools will make it easier for individuals and nonstate organizations to process and analyse this data. Earlier this year, it came to light that heat maps of jogging routes from runners wherein geolocating FitBits could be used to identify secret U.S. military and intelligence bases overseas.

Journalists quickly discovered that you could deanimize this data and actually identify specific users who had run routes as well as previous locations these users had visited.

This analysis was done manually but AI tools could make it easier to process this data at scale including linking it with other data sets such as social media profiles.

This kind of technology can make it easier for many to shine a light on state activity. It is not clear at this stage whether AI tools will benefit states or individuals more.

But it is clear that they are powerful and will be used by actors both to repress and enhance human rights.

I look forward to your questions. Thank you.

[The prepared statement of Paul Scharre follows]

PREPARED STATEMENT OF PAUL SCHARRE

May 22, 2018

Testimony before the Tom Lantos Human Rights Commission

Artificial Intelligence: The Consequences for Human Rights

Paul Scharre, Senior Fellow and Director
Technology and National Security
Center for a New American Security

Chairman Hultgren, Chairman McGovern, and distinguished members, thank you for inviting me to testify today.

Recent years have seen rapid advances in artificial intelligence and machine learning. AI tools are now coming out of research labs and into the real world, and are reshaping a variety of industries – medicine, transportation, finance, cybersecurity, and more. AI will similarly have important applications to human rights.

Artificial intelligence is a general-purpose enabling technology, much like electricity, computers, or networks. AI will be used by state and non-state actors for a variety of purposes, some of which will no doubt include suppressing human rights. Other uses may help to enhance human rights or fight against repressive regimes. It is not my intention today to estimate what the net effect of AI technology will be for human rights. Rather, I would like to walk through some features of the technology and some potential use cases to help illustrate some of the possibilities as AI technology becomes more widely used.

Applications of Artificial Intelligence

AI tools can be used for a variety of applications. Some examples include:

- **Data classification**, such as identifying images, classifying song genres, or arriving at medical diagnoses.[1] Given a sufficiently large set of training data, algorithms can be trained to classify data extremely accurately, often better than humans.
- **Anomaly detection**, such as finding fraudulent financial transactions or new forms of malware.[2] Traditional methods of anomaly detection require looking for known signatures. However, new AI tools can find anomalies whose signatures are not yet known by analyzing routine patterns of data and then identifying new data that is outside the norm. These systems can be used to monitor large data streams, such as financial transactions, at scale and in real-time in ways that would not be feasible for humans.

- **Prediction**, such as making statistical predictions about future behavior based on large datasets. Systems of this type are already widely used commercially, such as recommendation algorithms in Netflix and Amazon and search engine auto-fills. Other uses raise difficult ethical issues, such as predictive policing or predicting patient longevity in end-of-life care.[3]
- **Optimization,** such as improving performance and efficiency in industrial systems. Given a known goal, such as saving energy or reducing costs, AI systems can often find novel solutions to problems.[4]

AI and Threats to Human Rights

In the hands of a repressive state with access to large datasets about its population, these tools could be used to further increase state control. Automated facial recognition technology combined with security cameras could make *1984*-style continuous monitoring feasible in metropolitan areas. Combined with other readily available digital data collected through computers and smartphones, AI tools could be used to comprehensively monitor a person's behavior, communications, likes, and desires at a scale not even Orwell could have imagined.

People living in the digital age create a cornucopia of data: smartphone geolocation, browser history, web search history, online purchases, contacts, social media engagements, email and text message content, telephone calls, and more. Whoever has access to this data has tremendous insight not only into a person's past, but also the ability to predict their future behavior. Without AI tools, though, it is hopelessly impractical to manage this data at scale.

AI makes much of this data more discoverable through data classification tools that can recognize faces, identify human emotions, translate voice to text, translate languages, and process language. AI tools also make it feasible to analyze and process this data at scale. This means that the kind of intrusive monitoring that would in the past have been extremely time-consuming and resource-intensive can now be done quickly and at scale, allowing far more extensive and intrusive monitoring of a population.

Moreover, large datasets can be aggregated to generate statistically valid predictions. By learning from data across an entire population and then applying this to readily available information about an individual, AI tools could be used to make predictions about that individual's preferences or behavior – political, financial, sexual, or other. AI tools could be used to not only monitor a population, but predictively crack down on would-be dissidents.

AI to Assist Human Rights

At the same time, there are a number of features of AI tools that would make them powerful allies for those fighting repressive regimes. AI systems embed expertise within the software itself, lowering the bar the skills needed for a given capability. One does not need to spend years learning chess anymore to play at the level of a grandmaster; one can merely download a chess app for free. Similarly, AI systems will put greater abilities in the hands of non-state groups and individuals. Smartphones already turn surveillance tools back against the state, allowing citizens to record abuses

by authorities. AI tools such as embedded object recognition or facial recognition in the hands of everyday citizens could make it even easier to identify abusers and hold perpetrators to account.

A core feature of information technology is that it renders the costs of copying and transmitting information close to zero. One of the consequences of this is that it is difficult to keep information secret. While this is true for personal information, it is also true for state secrets. Individuals have accessed and released large tranches of government secrets on a scale that was impossible in the pre-digital era. The ease with which information freely flows in the digital age is a hindrance to repressive regimes that thrive on secrecy.

AI tools will make it easier for individuals and non-state organizations to process and analyze this data. In January 2018, a student at the Australian National University pointed out that "heat maps" of jogging routes from runners wearing geo-locating Fitbits could be used to identify secret military and intelligence bases overseas.[5] Journalists quickly discovered that they could de-anonymize the data and actually identify specific users who had run routes as well as previous locations they had visited.[6] This analysis was done manually, but AI tools could make it easier to process this data at scale, including linking it with other datasets such as social media profiles.

Embedding expertise within the software allows for greater automation, which can expand the scale at which smaller groups can achieve effects. For example, a few individuals have been able to cause significant internet disruption for short periods of time using botnets to infect Internet of Things (IoT) devices and launch distributed denial of service (DDoS) attacks.[7] Automation may allow small groups to achieve outsize effects, which levels the playing field against powerful actors and may be helpful in combatting repressive states.

Conclusion

A key question for any new technology is whether it concentrates power in the hands of a few or democratizes power towards the many. AI has features of both. At present, large datasets are needed to train AI systems. Additionally, the most cutting-edge advances in AI require significant computing resources.[8] At the same time, many AI tools are freely available for download online,[9] and much data is openly available. Artificial intelligence will enable actors who both seek to enhance human rights and those who aim to repress them.

CNAS Funding

CNAS is a national security research and policy institution committed to the highest standards of organizational, intellectual, and personal integrity. The Center retains sole editorial control over its ideas, projects, and productions, and the content of its publications reflects only the views of their authors. In keeping with its mission and values, CNAS does not engage in lobbying activity and complies fully with all applicable federal, state, and local laws. Accordingly, CNAS will not engage in any representation or advocacy on behalf of any entities or interests and, to the extent that the Center accepts funding from foreign sources, its activities will be limited to bona fide scholastic, academic, and research-related activities, consistent with applicable federal law. A full list of CNAS supporters and the center's funding guidelines can be found at https://www.cnas.org/support-cnas.

Mr. HULTGREN: Thank you, Mr. Scharre.

Professor Anderson.

STATEMENT OF KENNETH ANDERSON, PROFESSOR OF LAW, AMERICAN UNIVERSITY

Mr. ANDERSON: Thank you.

I am honored to appear before this commission.

Mr. HULTGREN: Can you make sure your microphone is on as well? I am sorry.

Mr. ANDERSON: Push. Yes. Thank you.

I am honored to appear before this commission in part because Representative Tom Lantos was someone that I knew early on in my career in the 1980s and he was a man of unshakeable integrity and commitment to these issues. So it's a pleasure for me to be able to appear here.

And I want to focus ultimately on the question of what I regard are important but limited tools that the U.S. government has in order to try and have an effect on where these technologies are used, how they're used, and ways in which to try and minimize their use in human rights abuse in particular.

And with specific reference to authoritarian regimes engaged in the internal repression of their own populations but specifically excluding China and the countries that are very large, who've got sophisticated programs, and really stand on their own, so it's in one sense a question of the follow-on from a place like China to other places in the world that are internally repressive.

I want to start by talking about the technology and the specific technology that we should be concerned with, I believe, for this kind of purpose.

The first distinction I would make is that for the purposes we are talking about here we are really talking about software and not physical robotics and thus I am not talking, for example, and I don't think any of us have been talking about automated weapons systems and those kinds of physical robots.

Instead, we are talking about software agents and within that category we are talking mostly today about another subdivision within that.

We tend to almost take for granted the extensive computerization and the programs that wind up operating hugely important automated systems in our lives, everything from getting the Social Security checks out to how telecommunication

switching devices work, and that stuff has been around and is developing and we somewhat tend to take it for granted for in relation to the new kid on the block that really what we are talking about is machine learning and much of controversy centering around a subset of that called deep learning, and these are essentially pattern recognition programs that are able to work their way through vast data sets and to be able to identify correlations that can be extracted from this.

Again, they have had some just enormously powerful successes in recent years including a role in the AlphaGo and Go game, which beat the experts using in part these kind of technologies.

But we also have become increasingly aware, particularly within the tech community, as they start looking at how these kinds of machine learning and deep learning programs go across this data without necessarily a completely fixed rule set for what it is they are looking to extract and how they are doing it, and have discovered that the learning process, which is essentially a process of reinforcement, can and, in important cases, does wind up reinforcing some of the things that we regard as socially abhorrent, illegal.

So machine learning programs that produce utterly racist results from data that just looking at it you wouldn't necessarily think that that's where it would wind up going. So there's been a much, much increased awareness, I would say, within the technology community that there is an important role for human beings to be looking from the standpoint of both common sense and ethical sense at what is generated out of these kinds of programs.

Nevertheless, those are the things that we are talking about because they have made the greatest strides particularly in facial recognition and software related to surveillance of enormous importance to such authoritarian regimes.

Now, one of the key points about those kinds of software programs such as facial recognition, surveillance, the automation of those kinds of processes, is that they are essentially going to by off-the-shelf programs with equally legitimate roles in policing, in national security, and a host of just things in ordinary commerce. So they're going to be there.

The problem is there's very little that separates what the program is used or legitimately from what that program is used for illegitimately in the way of human rights repression and therefore the question of what one does about it by policy isn't simply a matter of sticking a label on it and say this is a bad AI because it can be used illegitimately. Pretty much all of them can be.

The point that I would like to wind up on, however, goes the question of U.S. government policy and I've been sweating over this one because I think the options are actually limited at this point.

Traditional export control, licensing, that kind of control of it, I can't believe for a moment is actually going to be successful as something that is as universally available coming from places like China, Russia, and other places.

Moreover, regimes may be perfectly happy not to have a perfectly tuned system in the sense that we would regard it as required for our legitimate uses because they may not care about false positives, and I think Tom Lantos was aware of a case that came up in Guatemala.

I am not sure it was truly true but of the Guatemalan military apparatus deciding they wanted to eliminate someone. They didn't know which someone it was -- it was a common name-and went through the telephone directory eliminating name after name after name -- same name, different address.

If you don't care about the false positives then the off-the-shelf technology may be just fine for what it is you're looking to do.

Now, the one bright point about this, I think is that there is an opportunity I believe to work with the tech community in developing standards, broadly, under the name of ethical AI and AI ethics that will not solve the problem of illegitimate uses and the application of these technologies to evil ends, strictly speaking, but would have the ability to embed standards and norms that would make artificial intelligence programs more explainable and more transparent, among other things.

And if that became prominent across the commercial applications of this, I believe there's some case that it would spill over and bleed over into other kinds of applications.

That strikes me as one way in which one could look to go forward, but it is a very limited one.

And on that depressing note, I will close.

[The prepared statement of Kenneth Anderson follows]

PREPARED STATEMENT OF KENNETH ANDERSON

KA Draft Written Submission

May 22, 2018

US Congress Tom Lantos Human Rights Commission

Likely Impacts of Emerging Artificial Intelligence Software Agents

On Internal Human Rights Conditions in Authoritarian States

*

Artificial Intelligence: The Consequences for Human Rights

Hearing Before the Tom Lantos Human Rights Commission, United States Congress

Randy Hultgren, M.C. and James P. McGovern, M.C., Co-Chairs

2255 Rayburn House Office Building, Washington DC

May 22, 2018

Written Submission By

Kenneth Anderson, Professor of Law

Washington College of Law, American University

Washington DC

*

I. Summary

My thanks to the Tom Lantos Human Rights Commission for inviting me to make this submission on a question that is likely to take on increased importance over time: how emerging technologies in artificial intelligence software agents are likely to impact the internal human rights conditions of authoritarian regimes. I would like to preface my remarks below by saying that I had the privilege of meeting occasionally with Rep. Lantos in the 1980s when I worked as an NGO human rights lawyer. He was a person of great personal integrity and a deep, principled commitment to issues of human rights. It is an honor to be invited to make this submission to the Tom Lantos Human Rights Commission.

The key conclusions of my submission are that emerging applications of AI technologies will have important implications for the internal conditions of human rights in some, perhaps many, authoritarian countries – but that these applications are, today, still largely "emerging" rather than "emerged." The uses and misuses of these applications of AI by authoritarian states beyond a handful of technologically sophisticated pioneer states (China, most importantly) are largely still to come, and the contours of how they might impact particular societies much dependent on the specific characteristics of the applications, as well as the characteristics of the regime and society into which they are deployed, including the extent and sophistication of that society's digital infrastructure.

The policy implications for the United States government today are that it should be keenly observant of how such applications emerge, including their technological specifications, capabilities and limitations. In particular, it should absorb and take account of how these emerging AI applications are used and how they behave in both democratic and authoritarian societies, in order to understand ways in which these applications can be used and abused. This points toward taking account of what today is known as the field of "AI ethics" – interdisciplinary examination of the ways in which AI applications can be engineered and used in ethical ways, as well as ways in which, whether intentionally or unintentionally, these technologies wind up being used in unethical ways.

It seems unlikely to me that the US government will be able to prevent the spread of such AI software applications through long-standing methods of export controls, licenses, etc. There are too many potential producers of such applications; the US does not have a special lock on these

technologies, at least in their generic (and customizable) forms. It can assist US-based global technology companies in establishing industry standards in design, deployment, and use that might provide important normative markers, whether formal or informal, for acceptable uses of such technologies. It might be able to assist or encourage the development of AI applications – or applications drawing on other emerging technologies, such as distributed ledger or blockchain, cyber, or combinations of these – that might be of assistance to beleaguered human rights defenders at risk in authoritarian regimes, either in protecting themselves or in the work of gathering information on human rights abuses. But there are limits on how much the US government (or any government) is likely to be able to do to constrain the spread of these technologies or their illegitimate uses by authoritarian regimes.

II. AI Software Applications in Machine Learning

AI technologies and applications covers a vast range of possibilities, and it is important to understand certain key differences, including what technologies are clearly "emerging" and the nature of their likely capabilities and limitations. It is also essential to focus on "real," even if "emerging," technologies and applications of AI, rather than jumping to purely speculative possibilities for imaginary "AI."

The AI technologies and applications most relevant to the internal human rights conditions of authoritarian regimes are AI software agents – not physical, robotic machines. For that reason, as well as to avoid a range of very different normative and practical considerations, everything in this submission refers to pure software agents that run on computers, perhaps (and perhaps very likely) combined with cyber technologies – but not physical robots, such as autonomous weapon systems. As a general rule of thumb (and despite both the genuine successes but also hype surrounding self-driving vehicles), AI-enabled robotics is harder to do than pure AI software agents consisting purely of code; robotics involves sensors and motion/mobility in the physical world, and thus robotics requires whole fields of engineering not required by software programs alone.

A further narrowing of the field of AI to focus on the part most relevant to repression in authoritarian regimes today means drawing differences between "rules-based" AI and "machine learning" AI (in its several forms). What is normally understood as "computer programs" of the last few decades is computer software based around the execution of rules-based algorithms – the

rules of arithmetic, for example, in a calculator; we tend to forget that this is what the vast array of computerized functions in technologically advanced societies consists of, rather than the "emerging" AI techniques of machine learning (ML). For exactly the same reasons, however, that computer programs to automate such tasks as making social security payments to millions of individuals, or enabling telecommunications networks, or so many other things, allow society to work better, computer programs also exist that can automate such things as screening calls across a phone network for specific phone numbers believed to be used, for example, by dissidents or human rights defenders. These applications of rules-based AI computer technologies are so normal that we hardly think about them, but in fact form the large bulk of ways in which software can be used to repress in an authoritarian state.

The newer AI software applications comprising ML and its subcategories are today receiving most of the attention, but they are largely still "emerging"; have special social and technological requirements to be used effectively; and have uses (whether for good or bad) that are narrower than the existing range of applications of ordinary computerization. ML technologies are all about pattern recognition – various techniques for extracting patterns out of large quantities of data. The most important and most-discussed form of ML today is a type of so-called "artificial neural networks" (ANN) widely known as "Deep Learning" (DL). DL algorithms are largely at the heart of the current enthusiasm for AI technologies, and they are also at the heart of current controversies over AI applications and AI ethics. From the standpoint of both national security and human rights, DL has important implications because of the successes it has had in areas ranging from recent victories playing a strategy game such as Go to facial recognition software and related mass surveillance technologies.

DL successes have led to high hopes for the emergence of "predictive analytics" using "Big Data," among other things. In addition to applications such as AlphaGo or facial recognition software, DL has been used by private companies to create algorithms for, among other uses, purport to predict recidivism in the US criminal justice system (and already used in sentencing in some cases); individuals likely to be at risk from gang violence (used by some American police departments); buildings in a city likely to have a fire occur; and many more. Some of these algorithms work better than other prediction tools (including human experience and intuition); some of them don't; and with others, the lack of counterfactuals makes it difficult or impossible to know.

Indeed, a key and controversial aspect of DL algorithms is not just that they are hugely complex and opaque (true of code generally), but that they necessarily use probabilistic techniques that make it difficult to impossible to fully predict how the algorithm will behave ex ante or fully reconstruct how it did behave ex post. For this very reason, however, the "ethical AI" movement within the technology communities, at least in the open societies, has been pressing for new techniques and technological tools by which to evaluate how an algorithm acts, and to be able to assess whether a DL software program does what it is supposed to do and doesn't do what it's not supposed to do.

Important steps have been taken toward "Explainable AI," but there is still a distance to go in creating widely usable tools for "verification and validation, testing and evaluation" within the field of reliability engineering for these new forms of AI software. Moreover, one apparent finding in this field today is that, perhaps unsurprisingly, software can be made much more "explainable" – predictable up front or reconstructable afterwards – if it is *designed* to be explainable. This possibility of establishing norms for designing "Explainable AI" has implications for ways in which the US government, together with technology companies and governments of open societies, might be able to influence how DL algorithms with applications to surveillance, in legitimate national security ways or as tools of internal repression, can be generally engineered in accordance with industry common standards for transparency and explanation. It is by no means a "fix" to the human rights risks of DL algorithms, but it would matter if the routine, commercial or standard government, AI applications were built using widely accepted, verified and validated, "explainable" techniques – states could build their own without such features, or China or Russia or their companies might sell them, but it would help if there was a common commercial design norm favoring transparency.

The last important feature of ML and DL systems that matters to their use legitimately or illegitimately is that they are only as good as the datasets on which they "train." ML is "learning" because the algorithm is able to process a large number of examples relevant to the intended task – facial recognition, for example – from which it can learn correct and incorrect, within a probability range. In general, the datasets need to be very large in order to generate "accurate" learning, and smaller datasets can easily "teach" the machine algorithm systemically bad patterns – or simply produce results with many false positives or false negatives. Moreover, "datasets" actually means data that is digitized (while it's true that technological societies have large digitized data sets for some things, other societies do not, and much key information is not

captured digitally at all); accurate; and structured in such a way as to capture the intended features for analysis, and not inadvertently pushing the algorithm to learn unintended lessons.

The rash of straight-up racist ML outputs from otherwise non-racist datasets has alerted the technology community – less the application user community, so far – to ways in which ML programs can produce not merely incorrect, but socially abhorrent or illegal, results from datasets in which such outputs would not be obvious. There is almost certainly going to be a backlash, and perhaps regulation, in the US against untested and unverified algorithms that have negative impacts on individuals – lending decisions, for example, or criminal sentencing – in the US; Europe is already leading the way in terms of the regulation of the use of personal data and gradually emerging requirements of "Explainable AI."

Finally, as psychology professor and AI expert Gary Marcus has noted, DL algorithms perform far better at pattern recognition involving vast quantities of "primitive" data – pixels, for example, in facial recognition – than they do in higher level cognitive tasks. ML algorithms are about pattern extraction – correlations across large datasets – and not identifying causation or causes, or even the direction of causality in a correlation extracted from data. A ML learning algorithm developed in order to help predict who in an ICU was likely to die, using vast amounts of medical records, for example, achieved a remarkably good success rate in its predictions – so successful that its designers took a look inside the algorithm's black box. They discovered that the algorithm had focused with relentless literalness and no human common sense on a feature of the ICU medical records that had a box for the ICU physician to check, "Call hospital chaplain."

III. Likely Uses of AI Software Agents by Repressive Regimes

The uses of AI software agents, particularly ML and DL algorithms, by repressive, authoritarian regimes to monitor and control their own populations are likely to track the legitimate policing, intelligence, and national security uses of them. Essentially the same facial recognition software will be used – and available – to security services in open societies engaged in legitimate uses and in authoritarian societies where goal is to prevent dissent by identifying dissenters at an early stage. This means that attempts to restrict the technologies' use to "legitimate" purposes will always be somewhere between difficult to impossible.

Moreover, one of the engineering difficulties of ML algorithms is that testing and evaluating to ensure that, legitimate or not, the software identifies the correct persons and doesn't draw in

large numbers of false negatives or false positives is a difficult task. Off the shelf tech, even if sophisticated on its own terms, would certainly require extensive customization for any particular application in any particular society. If, however, you are an authoritarian regime that cares deeply about identifying dissenters, but doesn't much care if you identify too many people as dissenters who aren't – false positives – you might not be worried about off the shelf software that isn't customized with any sophistication.

The biggest limiting factors on the uses of such ML algorithms today in repressive societies outside the most important technological giants – China, e.g. – are likely two. One is simply that a society doesn't have enough digital infrastructure on which people routinely or necessarily depend, such as payment systems or banking, to use such digitally-based software on most people in the society. Additionally, with regards specifically to dataset driven ML algorithms, the digital infrastructure might not generate sufficiently large datasets that would run to the relevant information sought, e.g., facial recognition without widely used digital infrastructure ranging from ubiquitous video monitoring to Facebook users sufficient to make it likely that the relevant targets will be found. Thus, one reason why the use of ML algorithms specifically by many non-technologically sophisticated countries will not be immediate is that the population broadly doesn't provide the inputs for digitized databases. On the other hand, dissenters are often not the agricultural peasants, for example, but rather elites and those with access to the world through digital means. They do participate, and their tools of dissent are overwhelmingly likely to be digital. AI automation software combined with cyber monitoring tools can be a potent regime weapon to identify and surveil such dissenters – note, however, that these tools are already widely available, because they are not ML or DL algorithms, just ordinary computer programs monitoring digital communications.

[More TK in final version]

IV. Conclusions for US Government Policy

This submission has suggested that ML and DL tools that can be used for human rights violations and suppression of dissent within a particular authoritarian society are not yet widely available – but almost certainly will be. It would be a mistake to generalize from the example of China – gigantic and technologically sophisticated – to the many other authoritarian countries in the world. Those countries might not have the digital infrastructure at this point such that any

form of computerized surveillance would be effective – less so in the case of the cutting edge ML algorithms requiring large digital datasets. That said, digital sophistication will often be on the way in many authoritarian states – and the ability to monitor dissent by channeling members of society through digital tools under control or surveillance by an authoritarian government actually creates an incentive to invest in authoritarian-friendly versions of digital and cyber systems.

This submission has also argued that the appropriate role for the US government at this stage is to be sure to inform itself of possible ways in which such technologies could be abused – in part by paying close attention to the issues today of AI ethics of design. They are likely to also be ways in which such technologies are abused in authoritarian societies. The US government and governments of democratic countries might be able to work with their technology companies to come up with ways to limit the use of such technologies for illegitimate ends of human rights abuse and repression. This is likely easier said than done, however, because in many cases, the line between legitimate and illegitimate use of the technology will turn on the intent of the government in using it, to ends of legitimate policing and national security or illegitimate identification and suppression of dissenters in an authoritarian regime.

[This draft is not finalized; it will be extended, along with references, in its final version.]

END

Mr. HULTGREN: Thank you, Professor Anderson, and I do have more questions for all of you. So we'll transition into some questions, if that's all right.

I am going to start, Mr. Scharre, with you, if I might. As you know, China is already using artificial intelligence to abuse human rights.

Can you describe how the Chinese government is currently harnessing AI to the detriment of human rights and vulnerable populations, including their pernicious social credit system?

Mr. SCHARRE: Sure. There are certainly aspects where China is using already facial recognition technology. I think there has been quite a bit of discussion about the social credit system.

I want to separate for a moment kind of some of the hype surrounding media stories about it and what it actually does.

It's right now technologically relatively simple. It does not get to the place where they have access to large amounts of pooled personal data that is in fact aspirational as part of the system. But they do not currently have personal data included.

It's more of an ecosystem of existing lists that exist as well as there was for a period of time, although the contract has not been renewed, from the Chinese company Sesame -- credit scores.

It's predominantly oriented towards enforcement of existing Chinese laws and regulations so that there are consequences for people for, for example, not paying court fines or following through on existing regulatory or legal sort of judgments.

It exists as a set of lists -- there's a number of different lists. Right now, most of these, at least at the centralized national level, are binary lists so they're black lists that a person might get on.

The largest of these is the black list for the nonperformance of legally binding judgments. As of last year, there were 8.8 million people who were blacklisted.

Of course, in the total size of China it's not that massive but a significant number of people, and then there are some localized sort of regional lists, some of which actually at the local level do involve scoring for people based on performance.

Some people have scores that might make it positive points for things like taking care of the elderly or obeying the local laws and regulations and that might lose points for things like drunk driving or littering or jaywalking or other things.

The longer term aspirations of the system, you know, I think there is some cause for concern that it might start to lay the foundations for some of the more intrusive and comprehensive social monitoring that we talked about today.

Mr. HULTGREN: Thanks.

Professor Anderson, I wonder -- you mentioned a lot of the challenges with these issues and this trouble that we've got ahead of us.

I wonder, does the borderless nature of modern computing make it impossible to create effective international regulations that would prevent the misuse of AI to violate international human rights?

What's your thoughts on that? How would we even go about this, because it is not hindered by borders?

Mr. ANDERSON: I think the fact that it is borderless and the fact that many of these technologies are joined by hip with cyber technologies, one kind or another, make the ability to stop it at the border make it difficult and, ironically, the place that has got the greatest likelihood of being able to stop it at the border will be China when it comes to dealing with incoming stuff from the outside world.

I would actually flag in this regard something that has been discussed as the entire sort of question about Russian interference in the election and all that stuff has come up, and I am referring to Vladimir Putin's overture to the U.S. government -- that perhaps we ought to have a sort of declared noninterference policy which I think raises enormous questions related to this kind of AI and related to the connection to cyber in particular because it winds up essentially saying we are not going to engage in democracy promotion kind of efforts across borders.

So there's -- and supporting civil society groups in, say, Russia society. So I think that there are enormous risks to go -- to be present going down that route.

Then, finally, I would say with regards to international regulation in particular, I think that that's extraordinarily dangerous to contemplate because I don't at this point believe that there is a majority of the countries in the world that matter technologically that actually would favor the kind of regime that we would be talking about.

Mr. HULTGREN: Interesting.

Mr. Goswami, if I could address to you, we've talked about some of the negatives -- keep talking about quite a bit of that too over the next few minutes.

But I also wondered if you could talk about how U.S. agencies maybe could use AI to promote human rights. Are U.S. companies doing enough with the technology

available to them including AI to ensure their overseas operations are not harming human rights?

Mr. GOSWAMI: The short answer is no, we are not doing enough and, yes, we should be doing more and we can be doing more.

I think to approach that question I would say that AI is the technology that's -- or technologies that's helping us analyse issues better and really dig down to the deepest data point to really understand where we should intervene, et cetera.

Behind all that is our ability and our political will to intervene and I think that's where we are lacking. We don't lack the technical capacity or capability. We are lacking in the political will.

I think there's a lot that the Department of Labor can continue to do. There's a lot that the USAID and the State Department, Department of Justice can continue to do to use these new data means and data collection and local means to see how workers are being treated in their supply chains -- in U.S. supply chains overseas.

And we have enough tools right now that exist even without AI to know that. We can survey workers through mobile phones. We can ingest an NGO data sheet or a survey for what that NGO does across the world within two minutes and have it on our desktop here in D.C.

The point is we have to act upon it and, you know, we have to bring it up at these hearings but we also have to put pressure on ANC heads to use that information to act upon it.

There are lots of -- a few different examples of how machine learning in particular has enabled supply chain managers to find data from various different sources to take a comprehensive look at how a supplier is operating -- whether they have legal issues, whether they have complaints made, et cetera, by workers against them.

There's really not that much compulsion that compels a supply chain manager to take that and do something about it whether it's getting their legal counsel involved, whether it's putting sanctions against that supplier, whether it's renegotiating their contract, et cetera, et cetera, and that's what we need more of and I think that's where the U.S. government can step in and compel companies to do more of that.

Mr. HULTGREN: Thank you.

Mr. Scharre, your colleague, Elsa Kania, also noted that China has ambitions to lead the world in AI by 2030. Does it make a difference for human rights outside of China if Chinese companies develop a lead in AI technologies? If so, why?

Mr. SCHARRE: Yeah, it absolutely does. I think it does in a number of ways. One is that the surest way to influence this technology is to be the global leader. The U.S. has really had this in current information technologies by being the first mover, by being the global leader in information technologies.

Inherent in most of the technology that's proliferated around the world are embedded American values, particular, in this case, the values of the engineers who develop them in terms of things like openness.

And whoever is the global leader in AI their implicit values will be embedded in some of this technology in terms of things like privacy, fairness, transparency, explainability.

It's certainly in America's interest to remain a global leader, not just for AI technology used in the United States but so that the AI technology used globally is influenced from sort of an American perspective.

I also think it's really critical because the actual instantiation of some of these technologies can often end up having, you know, back doors or surreptitious means of people spying and collecting information.

There was recently a recall within the U.S. Department of Defense of drones manufactured by China -- the Chinese company called DJI. The Defense Department did not get into a whole lot of detail about why they had ordered their forces to short of shelve these drones.

But there were a lot of public openly available reports by journalists that this company had been recording geolocation data and audio files from these drones. So I think, certainly, it's a major concern and the U.S. would be well served to stay a leader in this space.

Mr. HULTGREN: Thanks, Mr. Scharre.

Professor Anderson, you talked about really maybe the impossibility or even the reality that there shouldn't necessarily be international regulations right now on this. But I wonder what you do recommend -- what can be done to address the threats posed by human rights -- to human rights by the misuse of AI.

Is this a case where we should focus on proactively incorporating human rights prevention into technology when it's developed or are there more questions of developing counter technology to circumvent the negative effects of AI after they emerge? Any thoughts on that?

Mr. ANDERSON: Yes. I don't believe that formal exercises in international rulemaking will wind up being useful.

I believe they would be profoundly counterproductive for the reason that I mentioned, which is, basically, I don't think that the direction that the world is headed globally is one that favors the values that Paul Scharre has mentioned here.

I also wind up thinking that there are an awful lot of countries out there and the ones that we particularly have in mind here in this hearing where it's a feature, not a bug, if you have got a button you can flip that basically says now it's going to be nontransparent -- now it's going to be sort of limited within -- nobody's going to be able to look at it, et cetera, et cetera, et cetera, et cetera.

So I think that realistically it's not just on practical grounds but on moral grounds that I think it's a mistake to be looking at trying to talk about international regulation of this kind of thing.

And instead, the far more important approach is the one that the U.S. has had to date and it's really what you have said, namely, that it's remaining the leader in the field and to the extent possible allowing values that are important to Americans and many other liberal democracies across the globe to remain part of the design features of these things, and it sounds weird to be talking about a technology that I described as equally useable for legitimate as illegitimate uses.

But these fundamental design aspects -- explainability, transparency, ability to have some idea if it looks like it's going wrong -- all of these things are not necessarily things that you want if you're a repressive internal society.

And so I think that remaining the leader in that way is actually the single most important thing that could be done.

Mr. HULTGREN: Thanks, Professor.

Mr. Goswami, can AI -- we talked about turning on and off switches, basically, of features of these things. I wonder, and I assume it can, but can AI replicate human biases? How does this occur and can it contribute to discrimination and what might we be able to do about that?

Can you make sure your microphone is on? I am not sure it is. Is it on?

Mr. GOSWAMI: I think so.

Mr. HULTGREN: Okay. Thank you. Sorry.

Mr. GOSWAMI: Yes. AI –

Mr. HULTGREN: There it is.

Mr. GOSWAMI: -- can replicate human biases. Simply put, it's like humans. Machines learn through repetition and through the data that you feed into it.

So if you feed data that has some inherent biases into a machine, that's what it will replicate. For example, if you want your machine to identify things in an agricultural field and in those images only brown people -- people that look like me -- are in those fields, then the machine might assume that only brown people are farmers. Those -- that's a very simple way of saying that yes, biases can be replicated.

That can have grave concerns for human rights. For example, if decisions are being made in the criminal justice field or other sectors, that data that is included if it's biased can have negative repercussions for those people.

For example, I think the city of Chicago piloted a project to use AI to determine outcomes of cases whether people should bail or not and they also used AI to see which areas in the city should get more police power.

However, it turns out that the systems also pointed to African American people and African American neighborhoods should get more police presence because the data that was being adjusted to make then learn that was a result of some of the biased practices that we've seen and have been alleged as well.

So, yes, biases, you know, can be replicated by machines. I think it's important to -- when we think about well, what can be done about that, on any kind of AI decision making process, whether it's, you know, you're a client for a mortgage or a credit card online to an automated system that's telling you where to police better, we have to look at the outputs to see if -- are those outputs being discriminatory and then why, and then unpack what data is building up to those outputs and then very simply put, those who are designing systems and those who are operating AI systems should reflect the diversity that we have as a society as well so those are less likely that those biases will be replicated.

Mr. HULTGREN: Thank you.

Mr. Goswami, following up, what are some other examples of human rights organizations harnessing the power of AI to assist human rights defenders or promote human rights?

Mr. GOSWAMI: That's a good question and I think the answer to that is a bit of offense and bit of defense as well. You know, we've touched upon human rights defenders who could be coming under threat by repressive regimes or other actors and are often using simple technologies to encrypt their communications with one another or keep hidden from digital footprints so that it doesn't pop up in a repressive regime.

I will give one example. When I was at Amnesty International USA a few years ago, we had access to about 30 years of very meticulously collected human rights data -- very well organized sheets of information about different human rights incidents and risks that have popped up over the past 30 years.

We partners with Purdue University and a nonprofit called DataKind and we got some volunteer data scientists to pore through I think it was 11 million lines and -- 11 million lines of data that we coupled with our existing current data on human rights risks that were coming in to see if we could predict an outcome, so a human rights crisis. And the tests that are volunteer data scientists did with 80 percent or so accuracy they were able to predict a binary outcome.

You can imagine that that has a lot of potential for an organization -- a large human rights organization to look at data coming in and analyse that with past data to find patters and warn human rights defenders of impending risks or concentrate their resources into going after human rights hot spots, et cetera.

So there's a lot of potential for it. It is a matter of resources as well. A lot of human rights organizations, a lot of human rights defenders are operating in very repressive environments with very little resources and access to technology as well. I don't think we should be -- I don't think we should be distracted by the shiny object that is AI. There's also a lot of just everyday technologies such as secure communication, et cetera, that human rights defenders can use with the right resources.

Mr. HULTGREN: That's great. Thank you.

I think all of you have referenced, maybe, Mr. Scharre, especially in your answers, of how important it is for America to lead and I would maybe add on to that, to lead in AI ethically.

You know, so not only lead the technology but also, hopefully, have that ethical voice as we are moving forward on that.

I wonder any suggestions you would have of how the U.S. government can boost its AI industry and especially an ethical AI industry, and what I could be recommending to my colleagues of how do we best help that or push that.

Mr. SCHARRE: Yes. I think there are a number of tools that the government has at its disposal that we can -- we can generate as a society to help grapple with these things.

Certainly, one of them is sparking a public conversation about some of these tools as we begin to see uses in a variety of contexts.

There was just a news article today about facial recognition technology being used by Amazon, being used in police departments. So I think, you know, venues like this where members of Congress or others can highlight things, bring these issues to bear, is very valuable.

I think dealing with some of these concerns requires a cross-disciplinary conversation that brings together technology companies, policymakers, lawyers, ethicists, and others, and members from the general society, writ large.

I think there are probably things the U.S. government could do in terms of R&D development where there might be places where we say there's certain strong private sector advantages to spend money on artificial intelligence. Governments need to invest and move this forward.

But there might be narrow areas where we say it makes sense to have government funding and investment in certain particular aspects of AI, for example, say, in more explainable AI where things are maybe more robust and reliable. Issues involving AI safety might be places where it would be very valuable to do that.

I also think a general -- you know, the more that we can begin to have kind of this open dialogue with tech companies. There are certainly ways that Congress can do that. There are ways that the administration can do that.

I would like to see a national AI strategy. I was encouraged by a meeting -- an AI summit at the White House recently, bringing together tech companies. I would like to see more of a continuous dialogue so that we have these conversation up front as technologies are being rolled out, that at least tech companies acknowledge what some of the potentials for misuse are and we begin to discuss those collectively.

Mr. HULTGREN: Do you think some that -- you said, you know, there's already been some conversation there. But do you think the American AI industry is focusing enough on how important it is for them and for us to lead in this ethical technology advancement in AI?

Is -- are the American AI companies recognizing their role in that? I would absolutely agree that we have more of a responsibility I guess to encourage that and push that. How much are they doing that already?

Mr. SCHARRE: Yes, I think there's more that could be done. For example, I was disappointed by the demonstration at Google recently using an AI that technologically was very impressive in terms of engaging with a human in actual speech -- mostly, that there wasn't a discussion up front about some of these ethical issues. Now, Google backtracked very quickly afterwards within a few days, saying, well, we wouldn't use this to try to manipulate or deceive people -- we would put a disclaimer up front.

I think that's positive that they adapted to the public response but I think disappointed that we didn't see that from the outset from some of these companies -- that they weren't raising these issues internally from the beginning before they brought it out publicly.

Mr. HULTGREN: Thanks,

Yes, Professor Anderson.

Mr. ANDERSON: To that, the -- what's has just been said I think is of enormous importance and part of it is because it has to do with the culture of the tech industry itself, in American tech companies but also tech companies abroad.

And in the case of what we describe as AI ethics, there are kind of two layers to it. One would be stuff that you might just regard as good design ethics -- the ability to look inside the black box and see how it operates.

You can think of that as being AI ethics but, really, it's only AI ethics because it hasn't been done and there's been a sort of sense that it can't be done in various kinds of ways. But I think there are ways in which basic research could really go after that proposition.

With regards to the general public, I think that one of the most important things that government at all branches of government could do would be to de-enchant AI, demystify it, probably stop calling it artificial intelligence because it isn't -- you know, as you said at the very beginning, it's a bunch of different technologies that don't necessarily share that much in common with each other.

And in bringing it down to earth, talking about it in ways that make it concrete to people what the kinds of bad outcomes are, there is just something so mysterious to me about artificial intelligence, robotics, all these areas, that cause us to reach to these sort of either utopian fantasies on the one hand or dystopian sci-fi on the other, and getting it down to sort of brass tacks I think can go a long way to getting the public behind the idea that there are just concrete issues here which are going to have be dealt with by law and regulation.

And, finally, most concretely would be I think there has to be a discussion with the tech companies about their business in China and about it is that they are going to legitimately reconcile a very different system with things that we really object to with their internal laws but understanding that what those folks make and sell is going to go to all the other places that we are concerned about.

Mr. HULTGREN: Yes, it's a great point.

I think you touched on this but, you know, specifically with -- when there's state actors who are abusing -- using AI to abuse human rights, who do you think really does have the responsibility to take the lead in fighting against that?

Is it individual governments? Is it international organizations or the tech industry? Who do you think ought to do it and who maybe has the best potential have a positive impact on that?

Mr. ANDERSON: Were it state actors, I believe that it's the responsibility of governments to take the lead. I don't think that one can really expect the global tech company headquartered in the United States but with, you know, interests and shareholders and all of that stuff everywhere to take the lead and I also don't think that we ought to expect them to make value judgements.

I think we ought to expect them to carry out a certain bare minimum and I think we certainly ought to expect that within the context of the tech world itself that certain basics -- you don't like to your users about who's on the other end of the phone line. Those things I think one can expect the tech industry to do. But to make China policy I don't think one can or to make policy with regards to some other place I don't think one can.

And so I do think that it is government's responsibility, again, that follow the theme that I've been saying here. I don't think that international organizations are capable of addressing it.

They don't understand it and I think that their interest to this point are completely fragmented.

Mr. HULTGREN: As a law professor, do you think we need more laws here in America right now specifically to address or is it enough for us just to start the conversation and trying to have -- using our influence to impact China and other places? Or do you think we need legislation in place?

Mr. ANDERSON: I think, again, speaking as a law professor for myself only, I think that this is the wrong moment to be introducing laws. I think that this is the right moment for introducing regulations about self-driving cars and safety.

But I think that one should start with the obvious cases where the risks are completely evident to us on the roads and then adopt a very careful kind of go-slow approach to see where laws would actually be useful.

You know, domestically I think that that means to a large extent forbearing from bringing the full tools of law that could be brought bear on, specifically, the tech industry and the reason for that is that one would like ideally to see them internalize these kinds of

ethics that go with the liberal democratic society into how they approach the tech design itself.

And with regards to things that are overseas, I think that there will come a point in which we will be looking to impose sanctions on various people and not just the kind that we've had with Russian companies and actors or the Chinese PLA in relation to national security stuff.

And in those regards I think that we are going to have a difficult conversation, given the level of economic entanglement.

Mr. HULTGREN: I agree.

Mr. Goswami.

Mr. GOSWAMI: Just to kind of act the notion of doing these laws, I think, aside from that question, there are some existing instruments and principles that we should be abiding by also in this space.

For example, the U.N. guiding principles on business and human rights that were passed in 2011 amongst much fanfare and agreement in the global community, including business and governments around the world, especially the U.S.

One of the core principles of the UNGPs is this notion of access to remedy, and back then it was probably envisioned that if it's a brick and mortar operation or it's a company that's polluting a water system that a community engages in or relies upon, but we need to apply that also to the AI space as well.

If a computing system led to a violation of civil rights or harm that was conducted by a company, then the same principle of access to remedy applies.

A citizen should have the right to access a grievance mechanism where they can get remedy for how they were wronged, including in the AI space as well.

Mr. HULTGREN: Go ahead.

Mr. SCHARRE: If I may on that one point, I think we certainly don't want to, you know, overly strangle innovation in the United States. But I do think that there's one area where it's worth us beginning to have a conversation about a legal or regulatory framework and that has to do with using AI tools to impersonate humans.

It's something that I think, you know, would have sounded like science fiction a few years ago, maybe sounds still like science fiction today -- oh, that's something out in "Blade Runner" or something.

But it's clear that the technology makes it possible to do that today for things like, certainly, engagement on social media. There's a significant fraction of social media users -- estimates vary -- are twitter bots or other types of bots -- as well as in now things like voice.

And I think it's worth a conversation. There are many states where it's illegal to record a conversation with someone without two-party consent and it's worth having a conversation here about -- not about whether it makes sense to have AI tools that can emulate human speech. That's very valuable to have human-based interaction. But to have tools that would deceive humans and should there be some regulation there, if nothing else to level the playing field for companies that will act responsibly because they're concerned about their, one, reputation -- maybe they're concerned about ethics -- to level that playing field with them and others who might not.

Mr. HULTGREN: Great. I am going to ask each of you in just a moment, if you wouldn't mind just doing a final close and the challenge, I guess, is -- and you have already touched on it in your testimonies and also answered the questions -- but I guess just kind of wrapping it up in a package.

One of the things -- Professor Anderson, I appreciate you bringing up Tom Lantos. He's someone who's a hero to me as well, and just incredible mentor, but also the other part of this original team was John Porter, a congressman from Illinois, who was -- it was my understanding it was kind of his idea for this human rights commission and working with Tom Lantos.

So all that to be said, one of the things I love about this -- and I am sorry, my co-chairman, Jim McGovern, was not able to be with us today, because I appreciate him so much. I respect him so much.

I don't always agree with him, but that's okay. You know, that is a good thing to have that give and take as long as we have that basic respect for each other and we absolutely do have that, and I think that was laid out by Tom Lantos and John Porter and others who were following in the footsteps.

But as we saw from today with the busy day -- we have a lot of things going on -- one of our primary objectives with the Tom Lantos Human Rights Commission is to gather information and then to be able to get that to our colleagues of challenges of what ought we be doing and what ought we not be doing.

And so I guess that is what I would ask in summation of what I can bring, certainly, to my co-chairman but also to my other colleagues who are on the commission and those who are not on the commission but care about the rights of humans here and around the world and potential threats or opportunities to help them.

So if I could just ask you to kind of go down the line and maybe just a minute or two of summation of what I could bring, from your perspective, of the most important things that we, as members of Congress, ought to do or ought to be aware of.
Mr. Goswami. Well, thank you again for your leadership and for chairing this hearing and this topic.

I think, in summation, I am going to bring up a story from Illinois. I used to be a lobbyist for the Chicago Coalition for the Homeless and worked with you quite a bit when you were in the Springfield State House, and I remember -- it was probably 2007 or 2008 when my colleagues partnered with the Illinois Department of Human Services to do the first ever statewide study on youth homelessness.

And back then, they used the latest and greatest research technologies available to them to find out, you know, where homeless youth were around the state.

And it was such a detailed survey that I think you could enter in a zip code and find out how many homeless youth were there. It was very well done study.

And I think they also came up with some numbers that for every homeless youth -- for every 156 homeless youth there was only one shelter bed.

And we thought, because this was in partnership with the state, that just knowing that would galvanize us into action and reprioritize resources, and some things did change. But homeless youth -- homelessness amongst youth is still a major problem in Illinois.

What AI has done and what other technologies has done it is able to understand these issues better. It's able -- it's cheapened the way we can collect this kind of information.

We can really get to the pinpoint of what factors are leading to this problem. But at the end of the day, we have to act upon it. We have existing laws. We have existing provisions on the books that can help us act towards human rights just like we could have in Illinois as well, but we just need to do it.

So one thing I would impress upon you and your colleagues is more political will to use the tools that are at our disposal to promote civil and human rights and to rectify the wrongs that have happened.

Mr. HULTGREN: That's great. Thanks. Thanks, Mr. Goswami. Appreciate it.

Mr. Scharre.

Mr. SCHARRE: Thank you.

I do think that there are a number of steps the United States can do to help ensure that this technology is used in ways that are more likely to enhance human rights. We are at a place where there's a lot of uncertainty about how the technology will unfold or what it might be used for. In many ways it's like trying to go back a few decades and imagine what computers and networks and smart phones might be used for.

And we are not able to imagine all of those but there are concrete steps we could take today. One, as we discussed, is remaining a global leader in artificial intelligence -- things like investing in R&D, STEM education. A national strategy on AI would help. I think that shining a light on abuses that do happen in other countries would be very valuable. Keeping an eye on something like China's social credit system as it evolves and as they incorporate more personal data and how it's used.

Publicly discussing these issues in our society I think are really vital. Reasonable people might disagree on how these tools are used.

I think transparency in what the government is doing at a local, state, and federal level is really vital as well as what private companies are, of course, doing with tremendous amounts of personal data that they have.

As we discussed, I think there's more that the tech community could do to begin to think about ethics in this technology as they're developing it and before they roll it out. And there are may be places where, as it matures, there are specific countermeasures or tools that might be able to develop and put in the hands of people to fight against repressive regimes, which may be valuable, too.

Thank you.

Mr. HULTGREN: Thanks, Mr. Scharre.

Professor Anderson.

Mr. ANDERSON: I would echo everything that's been said earlier here and I guess I would particularly stress the need to remain a leader in the tech industry.

I don't know that that necessarily requires that we have a national strategy. I am not sure that industrial policy has worked out all that well for other countries that have decided this is what they're going to do but even if it's only informal some sense that we ought to be the leader and ought to enable the conditions by which to lead.

The second thing goes specifically to Congress. I think that it was of enormous importance that Mark Zuckerberg testified in Congress, and one can have different views about what the quality of that testimony was, but I think that is both a carrot and a stick here for Congress to play.

One is the carrot of essentially saying we are behind you -- we are going to be behind you in ways that allow the technology and innovation to move forward and we are not going to sort of smother the baby at birth.

The stick is that much -- surprising as it might find it on some days, Silicon Valley is part of the United States of America and it does have obligations here, and I think that it's sometimes necessary to make that clear to Silicon Valley and I don't think it needs to be done with sort of a heavy stick that essentially distances it from the rest of the society.

But I do believe that letting Silicon Valley know that specifically with regards to the place where it lives, if you're going to play in society with big data, with things that affect every single one of us who's got an Amazon account, you got to play by society's rules and those rules are going to be made by society.

And then, finally, I guess I would say it looks to me unavoidable that there be an open conversation at this stage about what the obligations and the limits of obligations of global tech companies are to places where we have real objections to how they might be doing business.

I think we should be cooperating much more with the Europeans and I think they are ahead of us on certain things in relation to privacy and explainability in AI and I think we actually ought to be looking to them as an example of where we should be going in some of those things.

Mr. HULTGREN: Thank you, all. I really appreciate your time. I appreciate your expertise.

We need you. We'd ask if you'd stay in touch with us with suggestions or ideas of what we can do to move forward in this.

But I really do want to thank you for being a part of this.

And with that, the commission is adjourned. Thank you.

[Whereupon, at 4:02 p.m., the committee was adjourned.]

APPENDIX

MATERIAL SUBMITTED FOR THE HEARING RECORD

Tom Lantos Human Rights Commission Hearing

Hearing Notice

Artificial Intelligence: The Consequences for Human Rights

Tuesday, May 22, 2018

3:00 – 4:30 p.m.

2255 Rayburn House Office Building

Please join the Tom Lantos Human Rights Commission (TLHRC) for a **hearing** on the impact of artificial intelligence technology on global human rights.

Artificial intelligence (AI) refers to computerized systems that work and react in ways thought to require intelligence, such as solving complex problems in real-world situations, and that often involve machine learning. Often AI systems can make decisions without significant human oversight. AI is an emerging set of technologies that have yet to reach their full potential, yet it is already clear that AI technologies are capable of sifting through the vast amount of information available on the world wide web and social media, enabling their users to complete mammoth tasks that were previously impossible using standard computer programs.

However, for many observers, these technologies bear an uncomfortable resemblance to the surveillance described in George Orwell's *1984*. Concerns have been raised about whether these tools could be misused by malicious countries or individuals, permitting unfettered access to private information and improving their ability to target those that they consider "undesirables." There are signs that this is already happening in China.

This hearing will consider the impact of AI on human rights and human rights defenders. Witnesses will examine the potential harmful effects of AI, and discuss ways that the misuse of AI can be, if not prevented, mitigated. The positive potential of AI will also be discussed.

Panel I

- **Samir Goswami,** Consultant, 3rd Party LLC
- **Paul Scharre**, Senior Fellow and Director, Technology and National Security Program, Center for a New American Security
- **Kenneth Anderson**, Professor of Law, American University

The hearing is open to Members of Congress, congressional staff, the interested public, and the media. The hearing will be livestreamed via the Commission website, https://humanrightscommission.house.gov/news/watch-live and will also be available for viewing on the House Digital Channel service. For any questions, please contact Matthew Singer (for Mr. Hultgren) at 202-226-3989 or Matthew.Singer@mail.house.gov or Kimberly Stanton (for Mr. McGovern) at 202-225-3599 or Kimberly.Stanton@mail.house.gov.

Sincerely,

Randy Hultgren, M.C.
Co-Chair, TLHRC

James P. McGovern, M.C.
Co-Chair, TLHRC

PREPARED STATEMENT OF THE HONORABLE JAMES P. McGOVERN, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF MASSACHUSETTS AND CO-CHAIRMAN OF THE TOM LANTOS HUMAN RIGHTS COMMISSION

Tom Lantos Human Rights Commission Hearing

Artificial Intelligence: The Consequences for Human Rights

Tuesday, May 22, 2018
3:00 – 4:30 PM
2255 Rayburn House Office Building

Remarks for the Record

I thank Co-Chair Hultgren for convening this Tom Lantos Human Rights Commission hearing on artificial intelligence and its consequences for human rights, and extend my appreciation to our distinguished witnesses for their participation today. Regrettably, I am unable to be present due to Rules Committee and floor scheduling.

We as human beings, with our intelligence and heart, have the capacity to do great good in the world. But we are also clearly capable of inflicting great harm – not only one-on-one, but on masses of people, quickly or over extended periods of time.

The same social systems we create to help us order our lives can sometimes also be used to harm entire classes of people – as we saw with slavery in the United States or apartheid in South Africa. Our rules and norms can serve to protect us, or be manipulated to penalize and discriminate against us.

In the end, the difference between one outcome or the other depends on human agency. Do we harm the person who has angered us or do we step back? Do we attack those who are different from us, or do we recognize and embrace their humanity? Each of

us chooses every day to go in one direction or the other, as individuals and also as members and participants in the societies in which we live.

All of this was true before artificial intelligence appeared on the scene. But AI increases the stakes.

The technologies of artificial intelligence are one more example of the amazing capacity of human beings to learn about and transform our world. They are a testament to the incredible power of science, which has made possible so many improvements in the human condition.

But I am deeply worried that these technologies will be used – are already being used – in ways that make it easier for some human beings to make wrong decisions that harm others.

We already see the use of artificial intelligence technologies to facilitate social control. The Chinese are reportedly refining their capacity to use these technologies to surveil the Uyghur population in Xinjiang Province. Uyghurs are already so afraid that some outside of China have cut off contact with their relatives inside to protect them from retribution.

But of course the risks are not only with the Chinese. Here in the U.S. our own government is compiling biometric information on every person who crosses our borders. Cameras are everywhere. If I activate GPS on my phone, anyone can find me at any time.

Some of this is done in the name of security and some of it is about convenience. But in the end, already we are discovering that it is very hard to know the full extent of the data collected on us, much less to control what is done with it – or could be done with it, in the wrong hands. At a minimum, our rights to privacy, to freedom of expression and association, and to due process are all potentially at risk.

My second deep concern is with the use of artificial intelligence in weaponry – the development of "autonomous" weapons systems in which the machine or weapon or weapons program itself makes decisions regarding targets and kill zones.

In other words, killer robots.

These raise the grave human rights problem of who can be held accountable should a human rights crime be committed or civilians killed. They also raise the moral issue of taking a life based on a machine's software parameters.

It is encouraging to see that there is growing attention and debate over this issue, including discussion of the need for a new treaty banning the procurement of autonomous weapons. But we have a long way to go to protect ourselves and others from this moral scourge.

In the end, there are two great risks with artificial intelligence technologies. First, they concentrate knowledge and thus power in the hands of those who employ them. In places like China, that power is essentially unconstrained. The use of these technologies by authoritarian governments for surveillance, or to control access to information, goods and services, will make it very, very hard to organize opposition to any injustice.

Second, these technologies dehumanize decision-making. They make it all too easy to forget the human consequences of our decisions and our actions. Even with artificial intelligence, machines are not capable of compassion or empathy.

So I welcome this discussion today. We are in urgent need of ideas and recommendations that allow us to benefit from the advances of science without surrendering our humanity or our rights.

Thank you.

STATEMENT FOR THE RECORD OF AMNESTY INTERNATIONAL

May 21, 2018

Rep. Randall Hultgren
Co-Chair
Lantos Human Rights Commission
4150 O'Neill Office Building
200 C Street, SW
Washington, DC 20024

Rep. James McGovern
Co-Chair
Lantos Human Rights Commission
4150 O'Neill Office Building
200 C Street, SW
Washington, DC 20024

RE: **MAY 22 HEARING ON ARTIFICIAL INTELLIGENCE: THE CONSEQUENCES FOR HUMAN RIGHTS**

Dear Chairman Hultgren, Chairman McGovern, and Members of the Commission:

On behalf of Amnesty International[1] and our more than seven million members and supporters worldwide, we hereby submit this statement for the record. Amnesty International is an international human rights organization with major offices around the world, including the U.S. and the U.K.

Amnesty's Artificial Intelligence ("AI") and Human Rights Initiative

Amnesty's AI and Human Rights Initiative tackles human rights challenges posed by AI technologies. A core part of the initiative is the development of ethical principles for the development and use of AI. Amnesty International urges policymakers to enshrine such principles into existing human rights standards. Through our large network of human rights defenders and partner organizations worldwide, Amnesty International aims to facilitate dialogue with diverse global civil society voice about the ethics of AI, in order to ensure that the development of ethical and human rights principles for AI is guided by global human rights perspectives.

Building on our campaigning against the development of 'killer robots',

[1] Amnesty International was awarded the Nobel Peace Prize in 1977.

- AI systems collecting and processing vast amounts of personal data create new threats to rights, notably to personal privacy rights on both an individual and group level.
- A growing body of research demonstrates that AI systems are already contributing to discrimination – for example, in policing and criminal justice systems in the US. *The Toronto Declaration* underscores the risks to the right to equality and non-discrimination that are inherent to machine learning, and outlines means of protecting and promoting this right.[4]
- The impact of AI on policing and conflict could have extremely dangerous and irreversible implications on international human rights and humanitarian law.
- A lack of transparency and accountability in current systems denies those harmed by AI-informed decisions adequate visibility of harms and access to effective remedy.
- Innovation in AI technology is being led by powerful corporate actors and has rapidly advanced before appropriate state-based regulatory safeguards have been put in place.

Summary of recommendations

5. Amnesty International recommends that the US government:
   - Considers and acts to protect workers' rights and the right to work where AI technology is predicted to heavily impact employment practices, ensuring a gendered perspective.
   - Ensures that the rights of individuals, including privacy rights, are better protected through stronger data protection laws.
   - Introduces regulation to ensure that AI systems are audited effectively and system developers and users are held accountable, with clear processes of responsibility outlined prior to build and deployment.
   - Supports an international pre-emptive ban on the development, transfer, deployment and use of autonomous weapons systems.

---

[4] Amnesty International and Access Now led the drafting of *The Toronto Declaration on promoting the right to equality and non-discrimination in machine learning systems*, launched 17 May 2018. To date, over ten rights organisations have endorsed the Declaration, including Human Rights Watch and Wikimedia Foundation. Amnesty ultimately hopes that private sector actors and states will endorse the Declaration and acknowledge their existing commitments to the right to equality and non-discrimination.
https://www.amnesty.org/en/documents/pol30/8447/2018/en/

- Educates and informs citizens of their rights concerning privacy and data, including in automated decision-making.
- Invests in AI developments in the public sphere to foster AI technology and solutions for the public interest.

6. Amnesty International recommends that US-based companies:
   - Follow a human rights due diligence framework in order to ensure they have taken appropriate measures to avoid causing or contributing to human rights abuses through the use of AI systems.[5]
   - Take practical measures to promote AI systems that favour equity.

IMPACT OF AI ON EMPLOYMENT AND WORKERS' RIGHTS

7. Advanced AI systems will likely increase automation in the workplace. Technological advances and 'efficiency' savings will likely see machines replacing functions previously performed by humans in the workplace, as processes become part or fully automated.

8. The US government needs to approach the impact of technology on workers' rights from a gendered perspective. As more companies try to enforce a lower pay regime and weaker conditions of employment, women are highly likely to be disproportionately affected. The gig economy, if not properly regulated, risks lacking adequate protection for workers' rights thereby facilitating exploitation. At the same time, the expansion of automation is predicted to result in massive job losses, especially in the short-term, and especially at the expense of low-skilled positions, thereby risking further entrenching the social and economic marginalization of women.[6]

9. Authorities must act to regulate the gig economy in order to protect human rights. The growing spread of new forms of casual, on-demand work can prove beneficial, by allowing women to have more

---

[5] The responsibility of companies to respect human rights and carry out human rights due diligence is set out in the UN Guiding Principles on Business and Human Rights http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

[6] World Economic Forum, Towards A Reskilling Revolution, January 2018, p 13

flexibility with respect to their work life, whilst supplementing their income. However, when left unregulated this fragmentation and increased fluidity of the labour market can also pose serious risks for the socio-economic rights of women, as their protections are reduced and job and income security, discrimination, and exploitation worsen, thereby further entrenching unequal power relations in the work-place, in the family, and in society.

10. The US government needs to ensure that people can access their employment rights now and in the future, including:
    - Invest in training and reskilling programmes to help those whose jobs could be at risk of automation to stay employable, considering new skills that will be in demand in a tech-driven economy.[7]
    - Enable women to access decent work in the gig economy by implementing best practices such as parental leave, affordable and accessible care services (child, elder, disability); flexible working time arrangements (while respecting working time regulations); social security; basic infrastructure; discrimination protections; equal pay; safe working conditions and pension (particularly in the informal sector).
    - Prepare for an employment landscape that is radically altered by mass unemployment and fully considering the impact on state welfare and benefits systems. This may include exploring the viability and desirability of alternative income models like Universal Basic Income.[8]

PERSONAL DATA – PRIVACY AND PROFILING RISKS

11. Advancements in AI come hand-in-hand with the development of vast economies of personal data – raising concerns about privacy rights. AI systems are developed and trained using extremely large datasets. They are by and large designed to hone their function through continually processing new data – the larger quantities of

---

[7] The UK Parliament's House of Lords Select Committee on Artificial Intelligence recommended a significant government investment in skills and training to navigate the disruption in the jobs market. See Report of Session 2017–19, April 2018: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[8] For more on the human rights case for exploring Universal Basic Income, see report by Philip Alston, UN Special Rapporteur on Extreme Poverty and Human Rights, delivered to the UN Human Rights Council in June 2017: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G17/073/27/PDF/G1707327.pdf

relevant data that the system can access, the better. (For example, AI software in healthcare diagnostics will in theory perform better over time through collecting and processing live data from a wide source of patients to create more accurate diagnoses).

12. The right to privacy is hugely significant and yet widely abused by states through government mass surveillance programmes. Many governments, including the USA, have ultimately taken advantage of advances in technology to access and store private information on an unprecedented scale. The proliferation of AI systems creates the possibility for system owners to collect detailed and intimate personal information an individual level.

13. There are numerous risks associated with networked systems storing and processing such large amounts of personal data:

- Use of advanced AI software will dramatically increase the points of personal data collection in terms of both volume and detail. For example, facial-recognition and gait recognition technologies can easily capture and process detailed personal information on a previously unforeseen scale.
- The networking of interconnected systems – from the internet and telecoms, to systems and sensors in travel, health, logistics, traffic, electricity networks – allows the possibility for cross-referencing data that, if collected previously, used to be held in silos. Networked big data may be used to create intimate and precise personal profiles of individuals, a tactic already widely used for commercial advertising and political marketing during elections.[9] AI software makes profiling on such an intimate individual level much more accessible – with the potential for companies and governments to influence people to a greater degree than ever before, using highly personalised messaging across a range of platforms.
- Personal data is increasingly being used by systems to inform decision-making processes in all areas of our lives. There is potential for discrimination where information from one aspect of someone's life or previous behaviour is used to inform a decision or access to a service elsewhere. For example,

[9] http://www.bbc.co.uk/news/uk-39171324

insurance providers may use social media data to evaluate an insurance claim without the claimant's knowledge.[10]

AI SYSTEMS MAY PERPETUATE OR FACILITATE DISCRIMINATION

14. The adoption of AI and data-driven processes to aid governance and decision-making across many sectors of society has the potential to facilitate discrimination if proper oversights are not put in place. Working with a group of human rights and machine learning experts, Amnesty International and Access Now have launched *The Toronto Declaration,* which sets out the existing human rights obligations of states and responsibilities of private sector actors to protect the right to equality and non-discrimination in the context of machine learning, and outlines means of protecting these rights. The Declaration also highlights the need for systems (specifically machine learning systems, though the principles apply for related technology) to be visible, to allow individuals or groups means to challenge outcomes. Furthermore, the Declaration outlines existing obligations to ensure individuals and groups of people have access to effective remedy – a challenge for the current state and application of AI systems.

15. *The Toronto Declaration* was in part drafted in response to the serious problem with unconscious bias caused by the lack of diversity in the design of AI systems, which both states and private sector actors must address. The artificial intelligence and wider tech industry has seen a largely homogenous community power the creation and fostering of technology. The expertise and money for developing these systems is concentrated in a small pool of regions (US, North Europe, China). Systems are largely designed and deployed by a group of people with limited diversity in terms of race, culture, gender, caste, and socio-economic backgrounds.

16. As automated systems advance rapidly and are deployed across spheres with a high impact on human rights, there is an urgent need to put safeguards in place to mitigate the risks and guarantee accountability when abuses do occur. Scrutiny of such systems and

---

[10] Car insurance company Admiral last year attempted to use Facebook data to glean information that would inform insurance decisions: https://www.theverge.com/2016/11/2/13496316/facebook-blocks-car-insurer-from-using-user-data-to-set-insurance-rate

how they work as 'decision support' tools is difficult, given that these systems are usually proprietary. States must create means to regulate AI systems, particularly where they are used in public services, in order to ensure that rights are protected and people have access to effective remedy where rights are harmed.

17. The US Immigration & Customs Enforcement agency's proposed Extreme Vetting Initiative is a case in point.[11] The initiative sought to use automated decision-making, machine learning, and social media monitoring to assist in vetting of visa applicants and to generate leads for deportation. As set out, the program would have been both ineffective and discriminatory, proposing to evaluate whether an individual will become "a positively contributing member of society" or whether he or she "intends to commit criminal or terrorist attacks".[12] In a letter to the US government, 54 leading experts in machine learning and automated decision-making stated that "no computational methods can provide reliable or objective assessments of the traits that ICE seeks to measure" and that the proposed system would likely be inaccurate and biased.[13]

18. Another example is a highly-cited ProPublica investigation that found an algorithm used in the criminal justice systems of several US states to calculate a 'risk score' for prison inmates' likelihood of reoffending to be highly discriminatory.[14]

---

[11] In July 2017, ICE held an industry day in which it sought input from the private sector about an "overarching vetting contract that automates, centralizes and streamlines the current manual vetting process effort." ICE has since reportedly abandoned the proposal : https://www.washingtonpost.com/news/the-switch/wp/2018/05/17/ice-just-abandoned-its-dream-of-extreme-vetting-software-that-could-predict-whether-a-foreign-visitor-would-become-a-terrorist/?noredirect=on&utm_term=.6c56e8c72620

[12] Open letter to US Department of Homeland Security signed by 56 non-governmental organisations, November 2017:
https://www.brennancenter.org/sites/default/files/Coalition%20Letter%20to%20DHS%20Opposing%20the%20Extreme%20Vetting%20Initiative%20-%2011.15.17.pdf

[13] Letter to US Department of Homeland Security signed by 54 computer scientists, engineers, mathematicians, and other experts in the use of automated decision-making, November 2016:
https://www.brennancenter.org/sites/default/files/Technology%20Experts%20Letter%20to%20DHS%20Opposing%20the%20Extreme%20Vetting%20Initiative%20-%2011.15.17.pdf

[14] ProPublica, *Machine Bias*, May 2016

19. Predictive policing tools also carry a high risk of perpetuating discrimination. One research study from the Human Rights Data and Analysis Group (HRDAG) developed a replica of a predictive policing algorithmic programme that is used by police forces in numerous US states, and ran it as a simulation on crime data in Oakland.[15] They concluded that the programme reinforced existing racial discrimination within the police. This was because the system was built using already biased data that recorded higher crime rates in parts of the city with a higher concentration of black residents. The algorithm therefore predicted more crime in those areas, dispatching more frontline police officers, who subsequently made more arrests. The new data was fed back into the algorithm, reinforcing its decision-making process and creating a pernicious feedback loop that would contribute to over-policing of black neighbourhoods in Oakland.

20. Amnesty International has carried out research into the "Gangs Matrix" Database by the Metropolitan Police Service in London, UK, which uses an automated system to assign risk scores to individuals suspected of being 'gang members'.[16] The Matrix itself and the process for adding individuals to it, assigning 'risk scores' and sharing data with partner agencies appears to be ill-defined with few, if any, safeguards and little oversight. As a result, the matrix has taken on the form of digital profiling: 78% of individuals on the database are black, a number which is disproportionate both to the black population and the percentage of black people responsible for serious youth violence in London. In this context, the introduction of automated risk-scoring on top of an already deeply flawed data collection policy with no effective oversight and safeguards in place raises significant human rights concerns.

AUTONOMOUS WEAPONS SYSTEMS

21. Developments in AI over the last decade mean that it will be possible to develop and deploy fully autonomous weapons systems (AWS) which, once activated, can select, attack, kill and wound

---

[15] HRDAG, *To predict and serve?*, October 2016

[16] Amnesty International, *Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database*, May 2018

human targets, without effective and meaningful human control. Amnesty believes that these developments pose a very serious threat to human rights in the field of conflict and policing, and calls for an international pre-emptive ban on the development, transfer, deployment and use of autonomous weapons systems.

22. The use of AWS in law enforcement operations would be fundamentally incompatible with international human rights law, and would lead to unlawful killings, injuries and other violations of human rights. Effective policing is much more than just using force; it requires the uniquely human skills of empathy and negotiation, and an ability to assess and respond to often dynamic and unpredictable situations, which AWS would be incapable of. Decisions by law enforcement officers to use minimum force in specific situations require direct human judgement about the nature of the threat and meaningful control over any weapon.

23. Similarly, the use of lethal AWS would be incompatible with the three pillars of international humanitarian law; namely distinction, proportionality and taking reasonable precautions. AWS would lack the ability to analyse the intentions behind people's actions, or make complex decisions about the proportionality or necessity of an attack.

24. China, Israel, Russia, South Korea, the UK, and the USA, are among several states currently developing systems to give machines greater autonomy in combat. The history of weapons development suggests it is only a matter of time before this could spark another hi-tech arms race. This would cause these systems to proliferate widely, and end up in the arsenals of unscrupulous governments and eventually in the hands of non-state actors, including armed opposition groups and criminal gangs.

25. AWS also raises important issues related to transparency and accountability for human rights violations and individual criminal responsibility. Use of AWS would pose serious challenges to bringing accountability for crimes under international law. Under international human rights law, states have an obligation to investigate allegations of human rights violations and bring the perpetrators to justice as part of the right to an effective remedy – a

right which is applicable at all times.

26. In the case of lethal and less-lethal AWS, it is not possible to bring a machine to justice and no criminal sanctions could be levelled against it. However, actors involved in the programming, manufacture and deployment of AWS, as well as superior officers and political leaders, should be accountable for how AWS are used. But the nature of AWS is such that it would be impossible foresee or programme how an AWS will react in every given circumstance, given the countless situations it may face.

27. Furthermore, without effective human oversight, superior officers would not be in a position to prevent an AWS from committing unlawful acts, nor would they be able to reprimand it for misconduct. AWS, are by their very nature, autonomous agents that have no individual accountability. Deploying them in combat or for the use of force in civilian environments would be a perilous step for humanity, taking away one of the strongest deterrents against the unlawful use of violence.

TRANSPARENCY AND ACCOUNTABILITY

28. The inability to scrutinise the workings of all current deep learning systems (the 'black box phenomenon') creates a huge problem with trusting algorithmically-generated decisions.[17] Where AI systems deny someone their rights, understanding the steps taken to deliver that decision is crucial to deliver remedy and justice.

29. Provisions for accountability need to be considered before AI systems become widespread – practically, this may occur at multiple points, including in developing software, using training data responsibly, executing decisions. To what extent will any automated decision be able to be 'overridden', and by whom?

30. Restricting the use of deep learning systems in some cases may be required, where such systems make decisions that directly

---

[17] See for example, Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, 2015

impact individual rights. The US government should encourage the development of explainable AI systems, which would be more transparent and allow for effective remedies.[18] For example, a draft bill before New York City council advocates for transparency for all systems where algorithms are generating decisions in government services.[19]

38. Systems need transparency, good governance (including scrutiny of systems and data for potential bias), and accountability measures in place before they are rolled out into public use – especially where AI systems play a decisive and influential role in public services (policing, social care, welfare, state healthcare). It is vital that AI systems are not rolled out in areas of public life where they could discriminate or generate otherwise unfair decisions without the ability for interrogation and accountability.

39. There are also widely-applicable opportunities offered by AI systems in supply chain management, supported by blockchain technology for product identification, including provenance tracking and secure transfer of custody to provide transparency and accountability from product source to distribution. These include ensuring the tracking and movement of conflict-free goods and minerals.

40. Where there are potential adverse consequences for human rights, there must be higher transparency standards applied, with obligations both on the developers of the AI and the institutions using the AI system. This includes:
    - Detecting for and correcting for bias in design of the AI and in the data used.
    - Effective mechanisms to guarantee transparency and accountability in use, including regular audits to check for discriminatory decisions and access to remedy when individuals are harmed.

---

[18] An expert group of AI researchers has recommended that core public agencies, such as those responsible for criminal justice, healthcare, welfare, and education (e.g "high stakes" domains) should no longer use "black box" AI and algorithmic systems. See AI Now Institute, AI Now 2017 Report

[19] https://www.nytimes.com/2017/08/24/nyregion/showing-the-algorithms-behind-new-york-city-services.html

- Not using AI where there is a risk of harm and no effective means of accountability.

CORPORATE ACTORS

31. Government and civil society have struggled to keep up with the myriad of challenges to privacy and freedom of expression posed by developments in internet technologies: laws and public policies are still catching up with technologies that have been in wide use for years, if not decades. At the same time, there is a tension for policy-makers between the imperative to get to grips with and regulate the development and use of AI systems, and the appeal of these systems – which promise to 'modernize' and 'increase efficiency' across the public sector, while reducing cost. The overwhelming majority of AI systems are developed by private technology companies – systems which governments then may purchase to use in public services. As the uses of powerful AI technologies start to permeate all aspects of life, it is crucial that civil society and governments do not lag behind in responding to AI developments as they did with the development of the internet.

32. Amnesty is concerned that proprietary AI systems built by private actors will be in widespread use, including across the public sector, before human rights risks have been fully considered and addressed, and appropriate regulatory safeguards put in place. This presents a major barrier to ensuring transparency and accountability of such systems. Corporate actors themselves have a responsibility to respect human rights that exists independently from state's obligations. States need to ensure that positive developments in AI technologies, for example in healthcare, are not restricted by intellectual property practices.

CONCLUSION

33. To ensure personal data collection and use by AI systems does not impact negatively on the rights of people in the USA and around the world, the government must:
    - Create and uphold adequate regulation of private companies, including, for example, by mandating independent audits of AI systems where their use cases mean they can potentially have a significant impact on human rights.
    - Give greater powers to regulatory bodies that provide oversight

and accountability on the use of AI and big data, particularly where AI systems could adversely affect rights.

- Ensure that the rights of individuals, including privacy rights, are strengthened and upheld through stronger data protection laws, similar to the EU's General Data Protection Regulation (GDPR).
- Advocate for a pre-emptive international ban on the development, transfer, deployment and use of Autonomous Weapons Systems.[20]
- Ensure that AI systems in public service use are designed in a manner compatible with human rights standards, such as being non-discriminatory and providing means to pursue effective remedy.
- Invest in AI development in the public sphere to ensure development of AI technology and solutions for the public interest, and that it does not solely follow the commercial interests of private companies.
- Educate and inform citizens of their rights concerning privacy and data, including in automated decision-making.
- Restrict the use of AI systems that can't be interrogated in use cases where those systems make automated decisions that affect an individual or a groups' enjoyment of their human rights.

34. Companies and other private sector actors that develop and deploy AI systems and applications should:

- Follow a human rights due diligence framework to ensure they have taken appropriate measures to avoid causing or contributing to human rights abuses through the use of AI systems.
- Take practical measures to promote systems that favour equity, by investing in programmes that promote diversity of staff at development and deployment stage, and ensure that marginalised groups and individuals are not adversely affected by intentional or inadvertent discrimination.[21]

---

[20] Amnesty International urges the US government to engage in a comprehensive debate around the multiple challenges posed by AWS in order to develop and articulate a national policy on AWS (including less-lethal AWS) that takes full account of the state's obligations to respect and ensure international human rights law and international humanitarian law. This must be done in consultation with a broad range of stakeholders, including by meaningful and substantive engagement with non-governmental organizations and relevant experts, including AI and robotics experts and industry leaders.

[21] See *The Toronto Declaration* for suggested means of promoting equity and preventing discrimination in

For more information, please contact Joanne Lin, National Director for Advocacy and Government Affairs, at 202/509-8151 or jlin@aiusa.org.

Sincerely,

Joanne Lin
National Director
Advocacy and Government Affairs
Amnesty International USA

Anna Bacciarelli
Researcher/Advisor
Technology and Human Rights
Amnesty International

Joe Westby
Researcher/Advisor
Technology and Human Rights
Amnesty International

---

machine learning systems.

STATEMENT FOR THE RECORD OF FUTURE OF HUMANITY INSTITUTE, UNIVERSITY OF OXFORD

FUTURE OF HUMANITY INSTITUTE, UNIVERSITY OF OXFORD

AI: The Consequences for Human Rights

Initial findings on the expected future use of AI in authoritarian regimes

Paul de Font-Reaulx

House Foreign Affairs Committee
Tom Lantos Human Rights Commission Hearing:

Artificial Intelligence: The Consequences for Human Rights
Tuesday, May 22, 2018 – 3.00pm
2255 Rayburn House Office Building

This report is based on research conducted during a temporary stay at the Future of Humanity Institute. I thank my peers there for the invaluable input they have provided on this material, and particularly Jeff Ding for his remarkable help in bringing this report into existence. Any views expressed here are my own.

2. HOW CAN AI BE OF USE TO AUTHORITARIAN REGIMES?

Authoritarian regimes face a threat from a population that might mobilize into mass protests. If new technology would allow a regime to mitigate that threat, then we should expect the regime to make use of it. The recent developments in AI might provide a number of new such methods for authoritarian regimes to stabilize control. Below I consider four applications of AI which will plausibly be available to regimes within 10 years based on current research.

A) HIGH-PRECISION ALGORITHMS FOR IDENTIFYING DISSIDENTS

Advanced pattern-recognition systems developed in the last few years allows an actor to turn large amounts of data into useful information at an unprecedented scale. When individuals make decisions in their daily lives they provide different actors with such data, which includes for example revealing their location to various apps and their purchases to credit card providers. Sophisticated algorithms can be used to make reliable inferences from these data points to e.g. political inclinations.[1]

For instance, let us assume that someone who purchases only vegan food is statistically likelier to support gender equality policies. An actor that has data available on a person's purchases could then make a probabilistic claim about that person's political inclination. If they also have information that the person has been in the geographical vicinity of relevant political events, then they can infer with a high probability that the person has such views.[2]

An authoritarian regime with access to substantial data on its citizens could use such methods to infer highly relevant information about specific individuals. In particular it might allow the regime to gauge the degree to which a person supports the regime, and how much of a threat that specific individual poses. This makes repression less costly for a regime, as it can target a lower number of individuals and respond in an effective manner. It also makes it more attractive for a regime to rely on avoiding the spread of dissent as opposed to handling it once it does arise.

Knowledge that the regime has this capacity should dissuade individuals from engaging in regime-critical behavior. This is clearest in the case of protesters for instance. If you know that a regime can identify you with high reliability if you participate in a protest, and

---

[1] See e.g. Matz & Netzer (2017) and Sundsoy (2017).

[2] These techniques are already being put to use by US police for instance. The Chicago Police Department make use of algorithms to put together a 'Strategic Suspect List' of individuals who are statistically likely to be perpetrators or victims of gang shootings, and use this to inform them of the risk they are facing. Police in Los Angeles, New Orleans and New York have implemented similar techniques. (Ferguson 2017)

retroactively punish you if the protest is unsuccessful, then this would decrease your incentive to participate.

As will be discussed more below however, the effective implementation of such information-processing systems requires a high level of digital infrastructure and domestic organizational capacity which is unlikely to be found in most authoritarian states in the near future.

B) SOCIAL CREDIT SYSTEMS

Using the information-processing techniques considered above, a regime can systematize its response in the form of a social credit system or equivalent which rewards desirable behavior. This score can in turn be used as a supplement for financial credit for instance, or other benefits. Social credit systems are famously being explored by China today.[3]

Systematizing rewards and punishments in a way which is advantageous to a regime provides a cheap way of solidifying stability. It produces significant and consistent incentives for individuals to behave in a way commensurable with the benefits of the social credit system, which in turn dissuades regime-critical behavior which could constitute a threat. In the longer term there is some reason to believe that social credit systems could shape social norms to the benefit of a regime, which in turn might perpetuate long-term regime stability.

While China's Social Credit System has received much media coverage, there is reason to believe that this hype is somewhat overblown.[4] To the contrary of popular reports, it is unlikely that the Social Credit System will be fully implemented on a national obligatory basis by 2020. That does not mean however that this will not be the case in the future.

There are however other reasons to think that its spread could be limited. If a social credit system was highly centralized, then it would be fragile and constitute a security risk for the state due to its vulnerability to cyber-attacks. If it was made regional however, then the benefit in terms of stability would be contingent on regional elites, who might themselves constitute threats to the regime.

C) DISTORTION OF PUBLIC DISCOURSE

Advanced AI has allowed not just for the extraction of information from data, but also the creation of information made to confuse. In particular we have seen the development of 'bots' able to imitate humans in different contexts which are sometimes difficult to tell apart from

---

[3] For more on China's use of AI, see Ding's (2018) excellent overview.

[4] https://www.chinalawtranslate.com/seeing-chinese-social-credit-through-a-glass-darkly/?lang=en

real humans. As these technologies develop, we should expect it to become more difficult to distinguish bots from people in online interactions for instance.[5]

These bots can be used to shift the perception of public discourse in a way that is beneficial to the regime. By using bots to infiltrate forums online, e.g. comments on videos or news articles, a ruling coalition may use bots to create the impression that there is more widespread support for the regime than is actually the case. This could be used to dissuade would-be dissidents from taking action, giving the impression that they would be highly unpopular if they did. The other application of bots is to undermine the credibility of political opposition.[6]

The novelty of human-imitating bots is not the ability to influence sources of information for the population. This has often been done historically in the form of propaganda, and it is unclear to what degree it is effective, and to what degree it just causes the population to stop believing what they read. The difference with bots is that it might allow a regime to effectively infiltrate those forums which people might use as reference points to check whether national media is really credible, e.g. informal conversation online.

D) DRONES FOR ASSASSINATIONS

Military-style drones used to eliminate targets in warfare are already prevalent, and are currently being developed by numerous states. In the future we might also expect drones appropriate to assassinate targets to be developed. These could take the form of a cleaning robot made to explode once a target has been located,[7] or not to be bigger than an insect but able to inject a lethal agent into the target.[8]

Targeted assassination is intuitively an effective method of eliminating political opposition, and the threat of it a way to deter opposition. It is however not a novelty, and politically motivated assassinations have been conducted since ancient times. What is novel about employing drones is that identifying a perpetrator might be made more difficult due to untraceability. When there is ambiguity about who lies behind the death of some individual

---

[5] See e.g. Adams (2017) and Fariello (2017).

[6] Regimes like Russia already employ similar strategies today, by attempting to delegitimize opposition and portray them as extremists or criminals (Finkel & Brudny 2012)

[7] For this example, see Brundage et al. (2018, p. 27)

[8] For more on the development of coercive drones, see Scharre (forthcoming), Horowitz & Fuhrmann (2014) and Allen & Chan (2017).

the consequences are less severe in terms of e.g. popular dissatisfaction, as it is not clear who can be blamed.[9]

If untraceable drones become more easily accessible, then there is some risk that this will increase the prevalence of politically motivated killings by decreasing the risk involved in performing them. The effect can be strengthened if the assassination can be combined with a plausible cause of death, which might be provided by the information processing systems considered above (e.g. attributing death by a lethal agent to an existing heart disease).

It might however be that effective designs of drones with these capabilities will take many more years to produce, but their availability within 10 years cannot be ruled out. Furthermore, it might turn out that they provide little marginal benefit beyond currently existing methods of assassination, in which case we should not expect their availability to have significant implications.

3. REASONS TO DOUBT THAT AI WILL BE EFFECTIVELY IMPLEMENTED

In the sections above I have presented some of the probable effects of new technologies if they were to be effectively implemented by an authoritarian regime, and subsequently why such a regime would be interested in making use of them. Whether or not such technologies will be effectively implemented is currently an area of high uncertainty however, and there are a few reasons to doubt that they will.

THREATS FROM ELITES

One might think that the greatest threat against a ruling coalition is provided by unruly masses. The available data does not support this claim however. According to political scientist Milan Svolik 68% of authoritarian removals 1945-2008 have been by coup d'états lead by internal elites[10] (mainly the military)[11]. Because most of the technologies considered here seem more obviously applicable to minimizing the threat from popular protests, we should not expect such technologies to allow any given coalition to remain in power indeterminately, because the threat from other elites will remain.

---

[9] An example of this is the 2015 murder of Boris Nemtsov in Moscow. In this case some blame Chechnyan terrorists, which is considered a plausible explanation in the population. Other opposition politicians however attribute it to the Kremlin.

[10] Svolik (2012, p. 5), though Kendall-Taylor & Frantz (2014) present some evidence that protests have become more of a threat to authoritarian regimes today than during most of the 20th century.

[11] Svolik (2012, p. 149)

ORGANIZATIONAL CONSTRAINTS

Several of these technologies, particularly widely adopted information-processing and social credit systems, require a network of actors cooperating in cohesion. This might be unrealistic in most authoritarian states, where the dependence on relations with other elites is vital and often volatile.[12]

For instance, in order to have access to credit card information a regime would need to closely cooperate with a domestic credit card provider (or other actor with access to the data). Such private actors might however ally with a political opponent instead of the ruling coalition.[13]

Unless a ruling coalition can maintain effective control and the continued functioning of new technology, it will not be of much use, and the elite-dynamics of most authoritarian states might make this difficult.

INFRASTRUCTURAL CONSTRAINTS

Beyond the organizational requirements of effectively managing technology, there are also conditions of digital infrastructure which need to be satisfied. For instance, unless there is a functioning payment system with card in the state there obviously will not be much credit card information available to any domestic actor. A similar claim can be made about other sources of data.

For this reason we should not expect AI technology to receive widespread application in states that are unlikely to develop the necessary digital infrastructure to make effective use of it within the near-mid future. Subsequently I believe we should limit these discussions to states such as China or the Gulf States with that infrastructure, as opposed to including authoritarian states like Kyrgyzstan or Zimbabwe who are unlikely to develop it in the near future.[14]

---

[12] Bueno de Mesquita & Smith (2012) provide a good overview of these authoritarian dynamics.

[13] This happened for example 2004 in the Orange Revolution of Ukraine when Petro Poroshenko who owned the TV channel 5 Kanal allied with opposition leaders Viktor Yushchenko and Yulia Tymoshenko against the ruling coalition of Kuchma and Yanukovych.

[14] The judgments on infrastructure are based on the World Bank 2016 Logistics Performance Index, which can be found here:

https://lpi.worldbank.org/international/global?order=Infrastructure&sort=asc

4. WHAT ARE THE EXPECTED CONSEQUENCES IF THE TECHNOLOGY IS IMPLEMENTED?

While we should take the above points into account to avoid alarmism, they are not sufficient to rule out that authoritarian states will put these technologies to use in the future. In this section I note three areas where such developments could plausibly have implications relevant to human rights. I encourage the reader the remember that these developments are conditional on the technology being effectively implemented – which it might not be for the reasons considered in the previous section – and that they are in any case highly uncertain.

PROSPECTS OF POPULAR OPPOSITION

If these AI-technologies are effectively implemented by an authoritarian regime, then this will plausibly have a significant *negative* effect on the prospects of organizing popular opposition.

Regimes that have a substantial informational advantage on their citizens can both take action against dissidents before they constitute a serious threat and credibly threaten to take action against those who demonstrate dissent. This means that individuals who demonstrate high-probability dissent can be detained or otherwise debilitated early on to mitigate their influence, and should disincentivize individuals who otherwise would have considered opposing the regime.

This problem becomes most clear in the ability of the opposition to mobilize for protests. This is by itself difficult, as it requires coordination on the part of many people, and the incentives to participate are not strong when the protesting will get done whether or not you participate and put yourself in harm's way. If any brooding protest can be undercut before it spreads, then popular mobilization will be made even more difficult.

The regime might handle protests in even more subtle ways however. Suppose that the planners of the protest can be kept under strict surveillance (e.g. through the data they leave behind), distorting information released as to the purpose of the protest (e.g. to make it seem conducted by extremists), and more or less subtle punishments guaranteed for participation (e.g. identification by facial recognition causes significant drop in social credit). In that case a regime can even allow a protest to occur without it constituting any real threat to stability, and might even increase popular support for the regime among the general population.

In summary, unless counter-technology is developed to benefit political opposition in authoritarian regimes, I believe there is a high probability that popular opposition will be made more difficult as a consequence of AI-technology employed by the regime if effectively implemented.

PREVALENCE OF VIOLENCE

While these technological developments might increase the regime's power relative to its citizens, there is reason to believe that it would also cause a *decrease* in the use of violence by the regime.

Violent repression is often a costly measure by a regime. This is partly because the use of violence can act as a focal point for protests to mobilize beyond what can be controlled by a regime.[15] In other cases violence can lead to international repercussions or even intervention.[16] Another reason however is that repressing protests requires giving power to the coercive forces, particularly the military. Doing so however increases the risk that the ruling coalition will be subject to a military coup.[17]

For these reasons there is an incentive for a regime to rely on other means of mitigating the threat of popular protest than the use of violence. The technologies above mainly help a regime minimize the likelihood that uncontrolled dissent will occur at all, which is a cheaper way to stay in power than to expose itself to repercussions with the use of violence.

Therefore we can expect authoritarian regimes who effectively implement these AI-technologies to rely less on violence in order to stay in power. That is not the same as a decrease in repression however, which can take other forms. It should for instance make us expect an increase in subtle forms of repression on the basis of highly personal information, e.g. by blacklisting individuals from certain government-supplied services.

DEVELOPMENTS TOWARDS TOTALITARIAN DYNAMICS

Intuitively we might expect the implementation of these technologies to shift the dynamic between society and the state in ways that would make give it totalitarian traits.

Authoritarian regimes are non-ideological and function mainly as a form of organization between citizens and the regime. In totalitarian regimes on the other hand the state plays a more intimate role in people's lives, and by influencing norms and ideology they blur the distinction between state and society.[18]

---

[15] E.g. in Romania 1989 Nicolae Ceausescu was deposed by popular protests following a prior crackdown on protesters.

[16] This was for instance seen in the intervention in Libya 2011.

[17] This is what happened to Milton Obote of Uganda in 1971 when he gave increased powers to his army chief Idi Amin in order to suppress dissidents, who then performed a coup against him.

[18] See Linz (2000). Kazakhstan is a typical example of an authoritarian regime, while North Korea is a typical example of a totalitarian regime.

Given that the efficacy of some of the technologies above relies on promoting certain kinds of behavior, while discouraging others, we could intuitively expect their implementation to influence norms in society and how people behave in their every-day life. This is particularly the case when benefits and punishments can be based on highly personal choices and behaviors (e.g. choice of reading, conversation topics or drinking habits) that regimes did not have information on before, but now do due to the data that individuals inadvertently provide. Under such circumstances individuals would have an incentive to adapt these more personal choices and behaviors to avoid repercussions.

However, it is not clear that regimes would like these developments to occur, as the perceived infringement on privacy might cause dissatisfaction and dissent. Furthermore, other factors than technological development (e.g. leadership or culture) might be much more important to determine these dynamics.[19] Therefore, whether or not AI-technology has a significant effect on developments towards totalitarianism cannot be determined with any confidence now. By studying e.g. the Xinjiang region in China in the coming years however we should be able to discern the directions of such developments more clearly.

5. WHAT CAN THE UNITED STATES DO?

Any decision to take action should be made with the awareness that these developments are highly uncertain. These are many other factors which I and others might have failed to identify which could increase or decrease the impact of AI on authoritarian governance and human rights. In this report I have remarked on some reasons to be skeptical of a major impact. These reasons are not decisive however, and leave open the possibility that AI will have a significant impact possibly in the ways outlined above.

For this reason one major focus should be on understanding where these developments are heading. We have little available data at this point, but more will be available in the coming years. Understanding both what technological developments seem plausible based on the recent research, and how authoritarian regimes make use of the existing technology will be vital. Special attention should therefore be given to cases like Xinjiang in China where advanced AI technology is employed for the purpose of mitigating the risk of opposition.

If the US or other intentional actors would desire to take more direct action, it is not clear what would be effective. The necessary technology is often dual-use in nature, meaning that no international constraints can be put on it even in principle without also constraining other non-authoritarian applications of it. Constraining development is also difficult, as it can be

[19] The Soviet Union under Stalin for example was a typical example of a totalitarian regime, while under Khrushchev and Brezhnev it developed away from these totalitarian traits and became more of an authoritarian regime. Under the same period however there were many technological developments which would have facilitated totalitarian-style governance (e.g. the spread of the television).

done by private actors domestic to the authoritarian regime. The Social Credit System in China for instance is being developed by the Chinese company Sesame Credit. Those states that have the domestic capacity for development can in turn export this to other states.

Another alternative is to fund research into counter-technology to increase the power of political opposition in authoritarian regimes. This could be technology that provides decentralized communication that the regime is unable to influence, or software that provides misleading data to distort the information that the regime has on individuals. Whether such technology would be of more use to political opposition than the technologies considered here to a regime is currently an open question that likely deserves further research.

REFERENCES

Adams, Terrence. "AI-Powered Social Bots," 2017. https://arxiv.org/abs/1706.05143.

Allen, Greg, and Taniel Chan. "Artifical Intelligence and National Security." Belfer Center for Science and International Affairs, 2017.

Brundage et al., Miles. "Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation," 2018.

Bueno de Mesquita, Bruce, and Alastair Smith. *The Dictator's Handbook: Why Bad Behavior Is Almost Always Good Politics*. PublicAffairs, 2011.

Ding, Jeff. "Deciphering China's AI Dream", tech. rep. Future of Humanity Institute, University of Oxford, 2018.

Ferguson, Andrew Guthrie. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NYU Press, 2017.

Finkel, Evgeny, and Yitzhak M. Brudny. "Russia and the Colour Revolutions." *Democratization* 19, no. 1 (February 2012): 15–36. https://doi.org/10.1080/13510347.2012.641297.

Gusterson, Hugh. *Drone: Remote Control Warfare*. MIT Press, 2016.

Horowitz, Michael C, and Matthew Fuhrmann. "Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles," 2014, 30.

Kendall-Taylor, Andrea, and Erica Frantz. "How Autocracies Fall: The Washington Quarterly: Vol 37, No 1," 2014.

Linz, Juan J. *Totalitarian and Authoritarian Regimes*. Lynne Rienner Publishers, 2000.

Matz, Sandra C, and Oded Netzer. "Using Big Data as a Window into Consumers' Psychology." *Current Opinion in Behavioral Sciences* 18 (December 2017): 7–12.

Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. Norton, Forthcoming.

Schneier, Bruce. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. Norton, 2015.

Sundsøy, Pål. "Big Data for Social Sciences: Measuring Patterns of Human Behavior through Large-Scale Mobile Phone Data," 2017.

Svolik, Milan. *The Politics of Authoritarian Rule*. Cambridge University Press, 2012.