Comments from the AI Now Institute for the UN Special Rapporteur on Extreme Poverty and Human Rights on the visit to the United States
November 2017

As a part of his country visit to the United States, the Special Rapporteur has solicited input on the following topic:

"There is an increasing debate worldwide on the impact of new technologies on societies, including in the area of Artificial Intelligence, robotics, Big Data and algorithmic decision-making. How do these developments affect the human rights of those living in poverty in the United States? The Special Rapporteur is interested in learning how these technologies may affect civil and political rights as well as economic and social rights."

In 2016 and 2017, AI Now hosted symposia to bring experts, researchers, and the public together to discuss the broad implications of AI. In October of this year we released a report (provided alongside these comments) taking a deeper look at those implications, diving into how artificial intelligence and algorithmic decision-making has impacted issues surrounding **Labor and Automation, Bias and Inclusion, Ethics and Governance, and Rights and Liberties.**

In particular we believe the Special Rapporteur would be interested in our ten recommendations for how governments, industry, and the public should address AI systems. We share several recommendations that we feel are most relevant to the Special Rapporteur below. Also excerpted here is our report's study of how AI impacts issues of rights and liberties in particular.

Recommendations

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. This includes the unreviewed or unvalidated use of pre-trained models, AI systems licensed from third party vendors, and algorithmic processes created in-house. The use of such systems by public agencies raises serious due process concerns, and at a minimum they should be available for public auditing, testing, and review, and subject to accountability standards.

Expand AI bias research and mitigation strategies beyond a narrowly technical approach. Bias issues are long term and structural, and contending with them necessitates deep interdisciplinary research. Technical approaches that look for a one-time "fix" for fairness risk oversimplifying the complexity of social systems. Within each domain – such as education, healthcare or criminal justice – legacies of bias and movements toward equality have their own histories and practices. Legacies of bias cannot be "solved" without drawing on domain expertise. Addressing fairness meaningfully will require interdisciplinary collaboration and methods of listening across different disciplines.

Companies, universities, conferences and other stakeholders in the AI field should release data on the participation of women, minorities and other marginalized groups within AI research and development. Many now recognize that the current lack of diversity in AI is a serious issue, yet there is insufficiently granular data on the scope of the problem, which is needed to measure progress. Beyond this, we need a deeper assessment of workplace cultures in the technology industry, which requires going beyond simply hiring more women and minorities, toward building more genuinely inclusive workplaces.

On Rights and Liberties

Population Registries and Computing Power

In political contexts where minorities or opposition points of view are seen as threats to an imagined homogeneous "people," information technology has been used to monitor and control these segments of a population. Such techno-political projects often build on older colonial histories of censuses and population registries, as well as racialized modes of surveillance and control rooted in the Atlantic slave trade and the plantation system. In Dark Matters , Simone Browne connects this deep history of surveillance to contemporary biometric techniques of governing black bodies.

The Book of Life registry project in apartheid South Africa is a useful modern example. In that project, which ran from 1967 to 1983, IBM assisted South Africa in classifying its population by racial descent. This system was used to move all so-called 'non-white citizens' from their homes into segregated neighborhoods. The Book of Life was plagued by technical and operational problems and eventually abandoned. However, as Paul Edwards and Gabrielle Hecht note, "technopolitical projects do not need to fully achieve their technical goals in order to 'work' politically... The registries 'worked' to establish racialized personal identities as elements of governance." As Kate Crawford has recently argued, registries like the Book of Life were reinforcing a way of thinking that was itself autocratic.

More recent computerized registries like The National Security Entry-Exit Registration System (NSEERS) proliferated in the United States and among its allies following the attacks of September 11, 2011. NSEERS centralized documentation for non-citizens in the United States who hailed from a list of 25 predominantly Muslim countries that the Bush administration deemed dangerous. As with the Book of Life, NSEERS' effectiveness in its stated goal of stopping domestic terrorism was questionable, and it was dismantled in the final days of the Obama administration (although the data collected by the program still exists). Consistent with Edwards' and Hecht's analysis, NSEERS set into motion state projects of Muslim surveillance and deportation.

The history and political efficacy of registries exposes the urgent need for lines of research that can examine the way citizen registries work currently, enhanced by data mining and AI techniques, and how they may work in the future. Contemporary AI systems intensify these

longer-standing practices of surveillance and control. Such systems require the collection of massive amounts of data, which is now possible at large scale via the Internet and connected devices. When these practices are carried out by private enterprise in addition to states, as we will discuss in the next section, they introduce new forms of value extraction and population control unregulated and often unacknowledged by current legal frameworks.

Corporate and Government Entanglements

It remains critically important to understand the history of AI and its shifting relationship to the state. In the mid-twentieth century, advanced computing projects tended to be closely associated with the state, and especially the military agencies who funded their fundamental research and development. Although AI emerged from this context, its present is characterized by a more collaborative approach between state agencies and private corporations engaged in AI research and development. As Gary Marchant and Wendell Wallach argue, governance has expanded far beyond both governmental institutions and legal codes to include a wide range of industry standards and practices that will shape how AI systems are implemented.

Palantir—co-founded by Trump supporter and advisor Peter Thiel with seed money from the CIA's venture capital fund In-Q-Tel—typifies this dynamic. Gotham, Palantir's national security and government software, allows analysts to easily combine, query and visualize structured and unstructured data at large scales. AI can now be used in Palantir products for activities such as lead generation, including a bank's ability to identify anomalous credit card activity for fraud protection. More advanced capabilities are available to national security clients as well. How rights and liberties need to be understood and reconfigured in the face of opaque public-private AI systems is still an open question.

Immigration and law enforcement are critical within this debate. In the United States, Immigration and Customs Enforcement (ICE) is expanding its technological reach through tools like Investigative Case Management (ICM), a platform that allows agents to access a wide variety of previously separate databases, including information on a suspect's "schooling, family relationships, employment information, phone records, immigration history, foreign exchange program status, personal connections, biometric traits, criminal records and home and work addresses." This is another Palantir system, first procured by the Obama administration in 2014 and scheduled to become operational late in 2017.

Other law enforcement agencies are currently integrating AI and related algorithmic decision-support systems from the private sector into their existing arsenals. Axon (formerly Taser International) is a publicly traded maker of law enforcement products, including their famous electroshock weapon. The company has now shifted toward body camera technologies, recently offering them for free to any police department in the U.S. In 2017, Axon started an AI division following their acquisition of two machine vision companies. Among their goals is to more efficiently analyze the over 5.2 petabytes of data that they have already acquired from

their existing camera systems. Video expands Axon's existing Digital Evidence Management System, signaling a larger shift beyond machine learning and natural language processing of textual sources. Axon CEO Rick Smith has argued that the vast scale of existing law enforcement data could help drive research in machine vision as a whole: "We've got all of this law enforcement information with these videos, which is one of the richest treasure troves you could imagine for machine learning." There are real concerns about the forms of bias embedded in these data sets, and how they would subsequently function as training data for an AI system.

There are some who argue in favor of body camera and machine vision systems for supporting civil liberties, including enhanced law enforcement transparency and accountability. Axon promises that its AI techniques will reduce the time officers currently spend on report-writing and data entry. However, Axon's new focus on predictive methods of policing—inspired by Wal-Mart's and Google's embrace of deep learning to increase sales—raises new civil liberties concerns. Instead of purchasing patterns, these systems will be looking for much more vague, context-dependent targets, like "suspicious activity." Behind appearances of technical neutrality, these systems rely on deeply subjective assumptions about what constitutes suspicious behavior or who counts as a suspicious person.

Unsurprisingly, machine vision techniques may reproduce and present as objective existing forms of racial bias. Researchers affiliated with Google's Machine Intelligence Group and Columbia University make a compelling comparison between machine learning systems designed to predict criminality from facial photos and discredited theories of physiognomy—both of which problematically claim to be able to predict character or behavioral traits simply by examining physical features. More generally, Cathy O'Neil identifies the potential for advanced AI systems in law enforcement to create a "pernicious feedback loop"—if these systems are built on top of racially-biased policing practices, then their training data will reflect these existing biases, and integrate such bias into the logic of decision making and prediction.

Ethical questions of bias and accountability will become even more urgent in the context of rights and liberties as AI systems capable of violent force against humans are developed and deployed in law enforcement and military contexts. Robotic police officers, for example, recently debuted in Dubai. If these were to carry weapons, new questions would arise about how to determine when the use of force is appropriate. Drawing on analysis of the Black Lives Matter movement, Peter Asaro has pointed to difficult ethical issues involving how lethal autonomous weapons systems (LAWS) will detect threats or gestures of cooperation, especially involving vulnerable populations. He concludes that AI and robotics researchers should adopt ethical and legal standards that maintain human control and accountability over these systems.

Similar questions apply in the military use of LAWS. Heather Roff argues that fully autonomous systems would violate current legal definitions of war that require human judgment in the proportionate use of force, and guard against targeting of civilians. Furthermore, she argues that AI learning systems may make it difficult for commanders to even know how their weapons will

respond in battle situations. Given these legal, ethical and design concerns, both researchers call for strict limitations on the use of AI in weapons systems.

While predictive policing and the use of force have always been important issues, they take on new salience in populist or authoritarian contexts. As AI systems promise new forms of technical efficiency in the service of safety, we may need to confront a fundamental tension between technological efficiency and a commitment to ideals of justice.

AI and the Legal System

The legal system is the institution tasked with defending civil rights and liberties. Thus, there are two separate questions to consider regarding AI and the legal system: 1) Can the legal system serve the rights-protection functions it is expected to when an AI system produces an unfair result? And, 2) How and where (if at all) should the legal system incorporate AI?

Scholars like Kate Crawford of Microsoft Research and Jason Schultz of NYU have identified a series of conflicts between AI techniques and constitutional due process requirements, such as how AI techniques affect procedural considerations and equal justice under the law. The proliferation of predictive systems demands new regulatory techniques to protect legal rights. Danielle Citron and Frank Pasquale argue that safeguards to rights should be introduced at all stages of the implementation of an AI system, from safeguarding privacy rights in data collection to public audits of scoring systems that critically affect the public in areas like employment and healthcare.

In a similar vein, Andrew Selbst has argued that an impact assessment requirement can force those building and buying AI systems to make explicit the normative choices they are making before implementing them. And as Lilian Edwards and Michael Veale have pointed out, the new EU General Data Protection Regulation (GDPR) includes a requirement for data protection impact assessments, the import of which is unclear as yet. There is also a rapidly emerging scholarly debate about the value of requiring an explanation or interpretation of AI and machine learning systems as a regulatory technique to ensure individual rights, how to operationalize such a requirement, whether such a requirement presently exists under the GDPR and more generally how competing interpretations or explanations might be technically formulated and understood by different stakeholders.

The criminal justice system's implementation of risk assessment algorithms provides an example of the legal system's use of AI and its attendant risks. Proponents of risk-based sentencing argue that evidence-based machine learning techniques can be used in concert with the expertise of judges to improve the accuracy of prior statistical and actuarial methods for risk forecasting, such as regression analysis. Along these lines, a recent study by computer scientist Jon Kleinberg, Sendhil Mullainathan, and their co-authors showed that a predictive machine learning algorithm could be used by judges to reduce the number of defendants held in jail as they await trial by making more accurate predictions of future crimes.

While algorithmic decision-making tools show promise, many of these researchers caution against misleading performance measures for emerging AI-assisted legal techniques. For example, the value of recidivism as a means to evaluate the correctness of an algorithmically-assigned risk score is questionable because judges make decisions about risk in sentencing, which, in turn, influences recidivism – or, those assessed as “low risk” and subsequently released are the only ones who will have an opportunity to re-offend, making it difficult to measure the accuracy of such scoring. Meanwhile, Rebecca Wexler has documented the disturbing trend of trade secret doctrine being expressly adopted in courts to prevent criminal defendants from asserting their rights at trial.

Sandra Mayson has recently written on risk assessment in the bail reform movement. Well-intentioned proponents of bail reform argue that risk assessment can be used to spare poor, low-risk defendants from onerous bail requirements or pretrial incarceration. Such arguments tend to miss the potential of risk assessment to “legitimize and entrench” problematic reliance on statistical correlation, and to “[lend such assessments] the aura of scientific reliability.” Mayson argues that we also need to ask deeper questions about how pretrial restraints are justified in the first place. In other words, policymakers who hope to employ risk assessment in bail reform and pretrial forms of detention need to publicly specify what types of risks can justify these such restraints on liberty, as defendants receiving these scores have not been convicted of anything and these restraints are not imposed on dangerous individuals in the rest of society.

Separately, criminologist Richard Berk and his colleagues argue that there are intractable tradeoffs between accuracy and fairness—the occurrence of false positives and negatives—in populations where base rates (the percentage of a given population that fall into a specific category) vary between different social groups. Difficult decisions need to be made about how we value fairness and accuracy in risk assessment. It is not merely a technical problem, but one that involves important value judgments about how society should work. Left unchecked, the legal system is thus as susceptible to perpetuating AI-driven harm as any other institution.

Finally, machine learning and data analysis techniques are also being used to identify and explain the abuses of rights. Working with human rights advocates in Mexico, the Human Rights Data Analysis Group created a machine learning model that can help guide the search for mass graves.

AI and Privacy

AI challenges current understandings of privacy and strains the laws and regulations we have in place to protect personal information. Established approaches to privacy have become less and less effective because they are focused on previous metaphors of computing, ones where adversaries were primarily human. AI systems’ intelligence, as such, depends on ingesting as much training data as possible. This primary objective is adverse to the goals of privacy. AI thus

poses significant challenges to traditional efforts to minimize data collection and to reform government and industry surveillance practices.

Of course, privacy as a “right” has always been unevenly distributed. Rights-based discourses are regularly critiqued as being disproportionately beneficial to the privileged while leaving many vulnerable populations partially or entirely exposed. Yet what is different with AI and privacy is that while individualistic and rights-based conceptualizations of privacy remain important to some of the systems at work today, computational systems are now operating outside of the data collection metaphors that privacy law is built on. We are in new terrain, and one that 20th century models of privacy are not designed to contend with.

The expansion of AI into diverse realms like urban planning also raises privacy concerns over the deployment of IoT devices and sensors, arrayed throughout our daily lives, tracking human movements, preferences and environments. These devices and sensors collect the data AI requires to function in these realms. Not only does this expansion significantly increase the amount and type of data being gathered on individuals, it also raises significant questions around security and accuracy as IoT devices are notoriously insecure, and often difficult to update and maintain.

AI’s capacity for prediction and inference also adds to the set of privacy concerns. Much of the value that AI offers is the ability to predict or “imagine” information about individuals and groups that is otherwise difficult to collect, compute or distribute. As more AI systems are deployed and focus on ever-more granular levels of detail, such “predictive privacy harms” will become greater concerns, especially if there are few or no due process constraints on how such information impacts vulnerable individuals. Part of the promise of predictive techniques is to make accurate, often intimate deductions based on a seemingly-unrelated pieces of data or information, such as detecting substance abusers from Facebook posts, or identifying gang members based on Twitter data. Significant shifts are needed in the legal and regulatory approaches to privacy if they are to keep pace with the emerging capacities of AI systems.