13 March 2020

Mr. Dominic J. Mancini
Acting Administrator
Office of Information and Regulatory Affairs
Office of Budget and Management
725 17th St. NW
Washington, DC 20503

RE: Draft Memorandum to the Heads of Executive Departments and Agencies, Guidance for Regulation of Artificial Intelligence Applications

Mr. Maninci:

On behalf of Amnesty International[1] and our more than seven million members and supporters worldwide, we hereby submit this statement in response to the request for information on the "Draft Memorandum to the Heads of Executive Departments and Agencies, Guidance for Regulation of Artificial Intelligence Application" (henceforth "AI Memorandum"). Amnesty International ("Amnesty") is an international human rights organization with national Sections in more than 70 countries, including the United States (U.S.)

AMNESTY'S ARTIFICIAL INTELLIGENCE ("AI") AND HUMAN RIGHTS INITIATIVE

Amnesty's Artificial Intelligence ("AI") and Human Rights Initiative tackles human rights challenges posed by AI technologies. A core part of the initiative is the enforcement of human rights law to ensure that the development and use of AI does not undermine human rights standards, and instead strengthens such standards where possible.

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

OVERVIEW

1. AI has the potential to bring positive change for human rights. As noted in the *Guidance for Regulation of Artificial Intelligence Applications*, the deployment of AI technology "holds the promise to improve safety, fairness, welfare, transparency and other social goals." However, Amnesty believes there are key issues with AI systems that urgently need to be addressed to respect human rights.
2. Global innovation and development of AI technology is being substantially driven by major U.S.-based companies in Silicon Valley. The U.S. is one of a handful of countries that has experienced significant economic gain linked to the development of AI. The U.S. Government therefore has a critical role to play in adopting robust regulatory measures and accountability mechanisms to ensure that AI advances rather than undermines human rights. For the purpose of this paper, Amnesty adopts the definition of artificial intelligence as codified in Section 238(g) of the John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, 132 Stat. 1636, 1695 (August 13, 2018),[2] and also understands that the scope of the potential regulation covered by these regulations applies to "narrow" AI.
3. Amnesty's chief concerns with current AI systems are that:

[2] ARTIFICIAL INTELLIGENCE DEFINED.—In this section, the term ''artificial intelligence'' includes the following: (1) Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets. (2) An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action. (3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks. (4) A set of techniques, including machine learning, that is designed to approximate a cognitive task. (5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting

- AI technology is predicted to fuel significant changes to employment in the U.S., particularly through automation of jobs, which will require governmental action to protect U.S. workers' rights.
- AI systems collecting and processing vast amounts of personal data create new threats to human rights, notably to the right to privacy on both an individual and societal 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 U.S. *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.[3]
- 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

4. Amnesty recommends that Federal agencies considering regulations or policies related to AI applications:

---

[3] 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 on 17 May 2018. To date, over ten civil society 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 implement their existing obligations and responsibilities related to the right to equality and non-discrimination. https://www.amnesty.org/en/documents/pol30/8447/2018/en/

- Act to protect workers' rights and the right to work[4] where AI technology is predicted to heavily impact employment practices, ensuring a gendered perspective (i.e. ensuring that the rights of women workers are both understood and protected).
- Ensure any regulations of policies related to AI applications reflect the obligations of states under international human rights and the responsibilities of companies under the UN Guiding Principles on Business and Human Rights, including the rights to privacy and non-discrimination.
- Ensure that AI systems are regularly and effectively audited and system developers and users are held accountable for any adverse impacts on human rights, with clear processes of responsibility outlined prior to development and deployment.
- Ban the development, transfer, deployment and use of fully autonomous weapons systems.
- Ensure that affected citizens are informed of their rights concerning privacy and data, including in automated decision-making.
- Allow for affected citizens to challenge and appeal decisions made by automated systems.

ENCOURAGING INNOVATION AND GROWTH IN AI

5. According to EO 13859, "the policy of the United States Government [is] to sustain and enhance the scientific, technological, and economic leadership position of the United States in AI." The memorandum then states that "when deciding whether and how to regulate in an area that may affect AI applications, agencies should assess the effect of the potential regulation on AI innovation and growth. Agencies must avoid a precautionary approach that holds AI systems to such an impossibly high standard that society cannot enjoy their benefits."

[4] United Nations Office of the High Commissioner for Human Rights, *International Covenant on Economic, Social and Cultural Rights*, 1976, Article 6.

6. Amnesty strongly emphasizes that human rights must never be compromised. The goals set forth above do not excuse the U.S. Government from upholding its human rights obligations. We recommend that considerations of any regulatory impact on the protection of human rights be prioritized when deciding whether and how to regulate AI applications.

PUBLIC TRUST IN AI AND PUBLIC PARTICIPATION

7. We welcome the focus on both Public Trust in AI and Public Participation as key Principles for the Stewardship of AI Applications. However, ensuring Public Trust and effective Public Participation requires that Federal Agencies also provide an explicit right to remedy for individuals who are impacted by AI Applications.
8. Further, we welcome the specific guidance that "Agencies should provide ample opportunities for the public to provide information and participate in all stages of the rulemaking process." We suggest that Agencies consider direct outreach to affected and potentially affected communities, as well as working with and through existing community organizations.

RISK ASSESSMENT AND MANAGEMENT

9. The Memorandum states that "Regulatory and non-regulatory approaches to AI should be based on a consistent application of risk assessment and risk management across various agencies and various technologies…a risk-based approach should be used to determine which risks are acceptable and which risks present the possibility of unacceptable harm, or haram that has expected costs greater than expected benefits." We recommend that any risk assessment and management approach should explicitly include a human rights impact assessment.

BENEFITS AND COSTS OF AI

A. IMPACT OF AI ON EMPLOYMENT AND WORKERS' RIGHTS

10. 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.
11. The U.S. government needs to approach the impact of technology on workers' rights from a gendered perspective. 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.[5]
12. 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 employees to have more flexibility with respect to their work life, while supplementing their income. However, when left unregulated, this fragmentation and increased fluidity of the labor market can also pose serious risks, including 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.
13. The appropriate Federal agencies need to ensure that people can access their employment rights, including:

    - Implementing regulations that promote investments in training and reskilling programs 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.
    - Enabling men and women to access adequate work in the gig economy by implementing regulations that promote best practices such as parental leave, medical leave, affordable and accessible care services (child, elder, disability); flexible working time arrangements (while respecting working

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

time regulations); social security; basic infrastructure; discrimination protections; equal pay; safe working conditions and pension (particularly in the informal sector).

- Preparing for an employment landscape that is radically altered by the potential for increased unemployment and fully considering the impact on welfare and benefits systems.

B. PERSONAL DATA – PRIVACY AND PROFILING RISKS

14. 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 relevant data that the system can access, the better. For example, AI software in health care diagnostics will in theory perform better over time through collecting and processing data from a wide source of patients to create more accurate diagnoses.
15. The right to privacy is fundamental human right and yet widely abused through government mass surveillance programs. The U.S. Government can take 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 on an individual level.
16. 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 detailed personal profiles of individuals, a tactic already widely used for commercial advertising and political marketing during elections. 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 personalized messaging across a range of platforms.

- These capabilities mean there is a high risk that such systems could directly harm the rights to freedom of thought, conscience and religion and freedom of opinion and expression through their use of algorithmic systems.[6] Furthermore, they risk contributing to abuses of these rights by other actors who are able to access or utilise their models.
- Personal data is increasingly being used by systems to inform decision-making processes in all nearly areas of our lives. There is potential for discrimination where information from one aspect of someone's life or previous behavior is used to inform a decision or access to a service elsewhere.

FLEXIBILITY

17. The Memorandum states that "agencies should keep in mind international use of AI, ensuring that American companies are not disadvantaged by the United States' regulatory regime."
18. According to the UN Guiding Principles on Business and Human Rights, "Business enterprises should respect human rights. This means that they should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved."[7]
19. Efforts to ensure that "American companies are not disadvantaged by the United States regulatory regime" does not excuse the responsibility that these

[6] Rights guaranteed by UDHR Articles 18, 19; ICCPR Articles 18, 19.

[7] United Nations Office of the Human Commissioner for Human Rights, *Guiding Principles on Business and Human Rights,* 2011

companies face to respect human rights. Given the central role of many US companies in developing AI, we urge Agencies to assume an augmented role in terms of ensuring that the development of AI by American companies does not facilitate or otherwise undermine human rights.

EQUALITY AND NON-DISCRIMINATION

A. AI SYSTEMS MAY PERPETUATE OR FACILITATE DISCRIMINATION

20. 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 human rights protection and adequate oversight are not put in place. Working with a group of human rights and machine learning experts, Amnesty International and Access Now have published *The Toronto Declaration,*[8] 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 transparent and 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.
21. *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 AI and wider tech industry have seen a largely homogenous community power the creation and fostering of technology. The expertise and money for developing

[8] [8] Amnesty International and Access Now, *The Toronto Declaration on promoting the right to equality and non-discrimination in machine learning systems*, 2018, accessible at: https://www.amnesty.org/en/documents/pol30/8447/2018/en/

these systems is concentrated in a small pool of developed economies (e.g. in the U.S. and China). AI 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.

22. As AI 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 these risks and guarantee accountability when abuses do occur. Scrutiny of such systems and how they work as 'decision support' tools is difficult, given that these systems are usually proprietary. Agencies must regulate AI systems, particularly where they are used in public services, in order to ensure that human rights are protected, and people have access to effective remedy where rights are violated.
23. One 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.[9] According to Philip Alston, the UN Special Rapporteur on Extreme Poverty and Human Rights, "As humankind moves, perhaps inexorably, towards the digital welfare future it needs to alter course significantly and rapidly to avoid stumbling zombie-like into a digital welfare dystopia."[10]
24. 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 program that is used by police forces in numerous US states, and ran it as a simulation on crime data in Oakland.[11] They concluded that the program 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,

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

[10] Philip Alston, *Report of the Special Rapporteur on Extreme Poverty and Human Rights,* A/74/48037, 2019 accessible at: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25156

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

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 neighborhoods in Oakland.

SAFETY AND SECURITY

A. AUTONOMOUS WEAPONS SYSTEMS

25. 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 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.
26. 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.
27. 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 analyze the intentions behind people's actions or make complex decisions about the proportionality or necessity of an attack.
28. China, Israel, Russia, South Korea, the United Kingdom, and the US, 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.

29. 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.
30. 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, could be accountable for how AWS are used, though it is unclear who would be ultimately responsible. The nature of AWS is such that it would be impossible foresee or program how an AWS will react in every given circumstance, given the countless situations it may face.
31. 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 law enforcement environments would be a perilous step for humanity, taking away one of the strongest deterrents against the unlawful use of force.

DISCLOSURE, TRANSPARENCY AND ACCOUNTABILITY

32. The inability to scrutinize the workings of all current deep learning systems (the 'black box phenomenon') creates a huge problem with trusting

algorithmically-generated decisions.[12] Where AI systems deny someone their rights, understanding the steps taken to deliver that decision is crucial to ensure access to effective remedy.

33. Provisions for accountability need to be considered before AI systems become widespread. In practical terms, this may occur at multiple points during the life cycle of the system, including in developing software, using training data responsibly and executing decisions. It will also be important to consider the extent to which any automated decision may be 'overridden', and by whom.
34. Restricting the use of deep learning systems in some cases may be required, where such systems make decisions that have a significant impact on human rights. Federal agencies should encourage the development of explainable AI systems, which would be more transparent and allow for access to effective remedy when human rights are harmed.[13]
35. Systems must be transparent, good governance (including scrutiny of systems and data for potential bias), and accountability measures in place before they are used – especially where AI systems play a decisive and influential role in public services (policing, social care, welfare, 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.
36. 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 and correcting for bias in design of the AI and in the training data sets.

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

[13] 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

- Effective mechanisms to guarantee transparency and accountability in the use of AI systems, including regular audits to check for discriminatory decisions and access to effective remedy when individuals' human rights are harmed.
- Not using AI where there is a risk of human rights harm and no effective means of accountability.

CONCLUSION

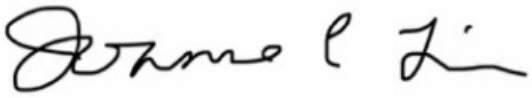
37. To ensure personal data collection and use by AI systems does not impact negatively on human rights, federal agencies must:

- Ensure that the human rights of individuals, including privacy rights, are protected.
- Create and uphold adequate regulation of private companies, including, for example, by mandating independent audits of AI systems where their use has the potential to have a significant impact on human rights.
- Put in place regulation, in meaningful consultation with independent technical experts and affected groups, to ensure oversight over the design, development and implementation of algorithmic systems to ensure companies are held legally accountable for human rights harms linked to such systems, including negative impacts resulting from the optimization decisions of such systems.
- Ban the development, transfer, deployment and use of fully Autonomous Weapons Systems.
- Ensure that AI systems used by government agencies are designed in a manner compatible with human rights standards, such as protecting the rights to privacy and non-discrimination and providing means to access effective remedy.
- Invest in AI development to ensure AI technologies and solutions have the objective of protection of human rights at their core, and that such technologies do not solely follow the commercial interests of private companies.

- Legally require technology companies to carry out human rights due diligence to identify and address human rights impacts related to their global operations, including risks and abuses linked to their algorithmic systems or arising from their business model as a whole.
- Invest in, encourage and promote the implementation of effective digital educational programs to ensure that individuals understand their rights, including their right to seek an effective remedy against any data protection, privacy, and other human rights abuses related to automated decision-making.
- Restrict the use of AI systems that cannot be interrogated where those systems make automated decisions that affect an individual or a group's enjoyment of their human rights.

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

Michael Kleinman
Director, Silicon Valley Initiative