White House Policy on AI

House Committee on Oversight and Accountability
Subcommittee on Cybersecurity, Information Technology, and Government Innovation

Testimony of Dr. Rumman Chowdhury
CEO and Co-Founder
Humane Intelligence

Chairwoman Mace, and esteemed members of the committee. It is an honor to speak with you today on the topic of President Biden's Executive Order on Artificial Intelligence as well as the recent Office of Management and Budget Implementation Guidance.

My name is Dr. Rumman Chowdhury and I am a data scientist and social scientist who has built AI and responsible AI for the past decade.[1]

Artificial intelligence is not inherently neutral, trustworthy, nor beneficial. This technology is not a mystery, it is not magic, and it is not alive. While it has immense capability, like many other high-potential technologies, it can also be used for harm by both malicious and well-intentioned actors. Concerted and directed effort is needed to ensure this technology is used to support and advance human interests.

[1] Dr. Chowdhury is the CEO and co-founder of Humane Intelligence, a tech nonprofit that creates methods of public evaluations of AI models, as well as a Responsible AI affiliate at Harvard's Berkman Klein Center for Internet and Society. Previously she built and led the practice at Accenture, the world's largest tech consulting firm, built the algorithmic auditing startup Parity AI, and led the AI Ethics, Transparency, and Accountability team at Twitter.

The executive order and the subsequent OMB guidance lay out an ambitious strategy for the accelerated responsible deployment of AI. I applaud the recognition that, in order for the US to remain an AI superpower, it must focus on safe, secure and trustworthy use. I offer the following recommendations to facilitate this goal:

1. First, the US must remain an active leader in the global AI landscape by funding targeted interventions in responsible AI for public and global use.
2. Second, in order to achieve the goals of Section 4 of the EO, support NIST.
3. Third, develop the independent community of algorithmic auditors by enabling secure model access, investing in education, and providing legal protection of structured public feedback methods, including red teaming and bias bounties.
4. Fourth, develop a minimum requirements standard to determine if AI adoption is necessary and appropriate for federal government use.

FIRST - Countries are moving quickly to establish global standards and best practices around responsible use. I arrived this morning from Singapore's AI for Public Good workshop. Their government gathered global AI experts to help co-create 10 projects to further AI for Public Good. Over the next year, the government will fund the development of these projects for open use and global benefit.

They are not alone. I've collaborated on efforts in London, Brussels, Paris, and Oslo where there is a similar investment in global responsible use best practices. The EU, for example, is actively exploring public and expert red teaming as an approach to compliance with the EU AI act. The OECD has already invested in the development of a global vulnerability database. The list goes on.

The US must continue to set global AI priorities, in alignment with Section 11 of the EO. I recommend that the government similarly invest in public interest projects to create methods, approaches and interventions for responsible use that are open access, publicly available, and a resource for individuals around the world. This could mean investment in digital public infrastructure, at-scale red teaming or funding talented US researchers to participate in global governance fora.

SECOND - Simply put, support NIST. Section 4 of the EO develops an ambitious strategy to leverage the institutional authority and capacity of NIST and expand their remit. I can think of no better team to execute on this plan. With a limited timeline and broad scope, they require significant funding and resources to deliver the global standard-setting quality that NIST is known for.

Similar institutes are funded accordingly. The UK AI Safety Institute has 100 million GBP earmarked for their endeavors. The Norwegian government has allocated a 1 billion Norwegian Kroner fund towards AI, of which there is a portion being utilized to ensure alignment with democratic values.

In addition, the proposed US AI Safety Institute must remain housed at NIST. As a scientific measurement body, they provide much-needed empirical evidence - data - to help us understand and prioritize how we address the risks and harms introduced by AI systems. They provide objective testing standards and a test environment that will help create consistent evaluations of AI system capabilities.

The US AISI must focus on a wide range of harms - societal impact, bias and discrimination, as well as broader considerations of future risks - in order to provide the full breadth of assurance that is needed to safely deploy. We already have significant evidence of AI systems in use today that infringe upon basic civil and human rights. These issues MUST be addressed immediately to ensure equitable adoption of AI.

THIRD - In June, I testified to the House Science, Space, and Technology Committee. At that time the concept of “red teaming” was known by the cybersecurity community and few others. Since then, it has become a topic of much consideration - mentioned fifteen times in the Executive Order alone. This is due in part to the White House’s support of the Generative AI Red teaming exercise this August, which was co-led by my organization, Humane Intelligence[2].

---

[2] https://www.whitehouse.gov/ostp/news-updates/2023/08/29/red-teaming-large-language-models-to-identify-novel-ai-risks/

We need to continue that momentum and enthusiasm. Methods of structured public feedback - bias bounties, expert red teaming, and public red teaming, are possible if there is talent and access[3].

I recommend provisions for secure model access for trusted third party entities to conduct algorithmic evaluations, as well as investment in algorithmic auditing practices as part of STEM education. Newly appointed CAIOs should engage with trusted organizations, including the US AISI and external independent organizations, to develop red teaming as part of a standard vendor procurement and project evaluation process within the federal government.

FOURTH - AI is often a hammer in search of a nail. We cannot assume that AI is always the best answer to a problem, as developers optimize for "efficiency" rather than effectiveness. We have already seen how the use of AI infringes upon civil rights via algorithmic discrimination in criminal justice, employment, banking, and more.

In 1971, the Supreme Court addressed a similar problem - unintentional employment discrimination introduced by aptitude tests. The Griggs vs. Duke Power Company[4] ruling established a minimum standard of adoption independent of intent, enabling the disparate impact requirement that is used widely in evaluating AI hiring systems today[5].

In order to achieve the goals of Section 7 & 8 in the Executive Order - advancing equity and civil rights, as well as protecting consumers - I recommend a similar approach for the use of AI in high risk situations in order to mitigate unintended consequences due to algorithmic bias. A minimum standards test could include the following:

---

[3] https://arxiv.org/abs/2311.14711
[4] https://supreme.justia.com/cases/federal/us/401/424/
[5] https://www.justice.gov/crt/fcs/T6Manual7 - Title VI Disparate Impact violation consists of three parts: (1) Disparate impact. Does the adverse effect of the policy or practice fall disproportionately on a race, color, or national origin group? (2) Justification: If so, does the record establish a substantial legitimate justification for the policy or practice? (3) Less discriminatory alternative. Is there an alternative that would achieve the same legitimate objective but with less of a discriminatory effect?

- Determination of whether an AI system performs better than equal investment in improving the current system
- Requiring alignment with impact metrics designed with NIST to measure the effectiveness of the system, not just performance efficiency
- Strategy to proactively identify and address real-world biases and adverse outcomes through in-context testing methods like expert or public red teaming

IN CONCLUSION, we must be circumspect on if, when, and how we adopt AI systems. This technology is meant to serve humanity and innovation is only possible if we are all able to reap the benefits. Thank you for your time.