SUBMITTED WRITTEN TESTIMONY FOR FULL COMMITTEE HEARING OF THE HOUSE OF REPRESENTATIVES COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

"ARTIFICIAL INTELLIGENCE: ADVANCING INNOVATION TOWARDS THE NATIONAL INTEREST"

JULY 22, 2023

My name is Dr. Rumman Chowdhury and I am a data scientist and social scientist who builds innovative sociotechnical solutions in trustworthy, safe, and responsible artificial intelligence (AI). In my career, I've helped address some of the biggest problems in AI ethics.

Seven years ago, amid growing concerns regarding the widespread use of AI, Accenture - the largest tech consulting firm in the world - tasked me with building a global practice in the nascent field of Responsible AI. At that time, the narrative around AI hype as well as AI disaster hype mirrored today's headlines - will AI take our jobs? Will AI end humanity? Will AI be like HAL or Terminator?

Reality is not a movie, and building solutions in Responsible AI meant grounding my work in the practical and tangible. During that time, I worked with global Fortune 500 leaders, and my team built the first enterprise algorithmic bias detection and mitigation tool. After my time at Accenture, I founded an algorithmic auditing startup, Parity AI, and was approached by Twitter, to build and lead a team focused on Machine Learning Ethics, Transparency and Accountability (META) as its Engineering Director.

During my time at Twitter, we pushed the boundaries of transparency and accountability - conducting the first ever industry algorithmic bias bounty, which was a public access challenge to identify problems in Twitter's algorithms.

In addition, I have long been a part of the vibrant ethical AI nonprofit and academic community: I am currently a Responsible AI Fellow at Harvard's Berkman Klein Center for Internet and Society, and co-founder of the nonprofit Humane Intelligence. I am also a co-creator of the largest public AI Red teaming competition which is supported by the White House[1].

As Congress considers AI regulation, innovation and governance, I offer the following considerations.

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.

My career in Responsible and Trustworthy AI can be described by my commitment to one word: governance. Governance is a spectrum ranging from norms and codes of conduct, to standards and benchmarks, to oversight and regulation. In order to remain competitive and innovative, the United States needs to intentionally focus on significant investment in all aspects of governance.

It is important to dispel the myth that "governance stifles innovation." This is not true. In my years of experience delivering industry solutions in Responsible AI, good governance practices have contributed

[1] https://arstechnica.com/information-technology/2023/05/white-house-challenges-hackers-to-break-top-ai-models-at-def-con-31/

to more innovative products. I use the phrase ‘brakes help you drive faster’ to explain this phenomenon - the ability to stop a car in dangerous situations enables us to feel comfortable driving at fast speeds. Governance *is* innovation.

This holds true for the current wave of AI. Recently a leaked Google memo declared ‘there is no moat’[2] - suggesting that AI will be unstoppable as open source capabilities meet and surpass closed models. There is also concern about the US remaining globally competitive if it is not investing in AI development at all costs. This is simply untrue. Building the most robust AI industry isn’t just about powerful models, processors and microchips - the real competitive advantage is trustworthiness.

To bolster and advanced trustworthiness and advances in AI innovation in the national interest, I offer four recommendations:

- First, support for AI model access to enable independent research and audit
- Second, investment in, and legal protections for, independent red teaming and ethical hacking
- Third, the development of a technology oversight body to supplement existing US Government efforts
- Fourth, development of and participation in a global AI governance organization

I. Data and Model Access and Transparency

Good governance starts with support for open access to data and models, to enable independent research and audit of AI systems. Purely closed, “black-box” AI systems that are unleashed on society are irresponsible. Due to the closed nature of these systems, impartial and objective external actors are unable to verify what the true impact of these models are - we can only observe anecdotally.

CEO’s of the most powerful AI companies will tell you that they spend significant sums to protect users, publish research, develop standards, and more. This is true - I was one of those people. My team and I held ourselves to the highest ethical standard, as do many of my colleagues who remain in these roles.

However - a well-developed **ecosystem** of governance also empowers individuals in academia, civil society, and government - trusted parties whose organizational missions are to inform and protect society. The European Union’s Digital Services Act, Article 40, creates such access for European Researchers. Similarly, the UK has announced that Google DeepMind and OpenAI will enable access to models for their government.[3] I recommend the United States match this innovative approach.

Tech company CEOs may push back on increased access due to concerns of intellectual property, security, and privacy. This is increasingly a concern of the past. Today, we have technologies that enable access to sensitive data without these issues. Privacy-enhancing technologies (or PETs) can and should be used to improve and mandate access to third-parties. I applaud the ongoing efforts of the US and UK Governments in conducting the PETs challenge at the Summit for Democracy, and encourage continued investment with the purpose of enabling oversight access[4].

[2] https://www.economist.com/leaders/2023/05/11/what-does-a-leaked-google-memo-reveal-about-the-future-of-ai

[3] https://www.politico.eu/article/openai-deepmind-will-open-up-models-to-uk-government/

[4] https://petsprizechallenges.com/

II. Investment in, and Legal Protections for Independent Auditing

Investing in AI Oversight Talent

New laws, such as the EU's Digital Services Act Article 34 and New York City's audit law for automated decision making in hiring (Local Law 1894-A[5]), are mandating third party algorithmic auditing. However, there is currently a workforce challenge in identifying sufficiently trained objective third-party algorithmic auditors. The few individuals who are skilled at these tasks are often hired by tech companies.

Alleviating this shortage requires two things: first, funding for independent groups to conduct red teaming and adversarial auditing, similar to 'ethical hacking' more common in traditional security fields, and second, legal protections so these individuals - operating in the public good - are not silenced with litigation.

Methods of incorporating public feedback - such as reinforcement learning with human feedback, red teaming and bias bounties - are a fairly new phenomenon in artificial intelligence. My colleagues at Google DeepMind, Open AI, Anthropic and more have engaged in this practice to refine their flagship AI models. But we need this practice to be supported by non-corporate actors as well.

My team at Twitter held the first Algorithmic Bias Bounty[6]. Last fall, I co-designed the first public bias bounty challenge[7]. Now, I am part of a group designing the largest-ever AI red teaming challenge.

"Red-teaming" is a process by which experts attempt to find vulnerabilities in an organization's systems to improve security and resilience. Invited third party experts are given special permission and access by AI companies to find flaws in their models. Traditionally, these practices happen behind closed doors, and public information sharing is at the company's discretion.

We will provide unprecedented access to most major open and closed source AI Language models. Thousands of individuals attending AI Village at DEFCON, the largest hacker conference in the world, will compete to identify how these models can intentionally or unintentionally produce harmful content. In addition, we are forging new territory with groups like MITRE on AI incident reporting, responsible disclosure of vulnerabilities, and responsible release of data.

This endeavor represents a collaborative, solutions driven initiative to create meaningful oversight of AI systems. Our efforts would not be possible without the support of the White House Office of Science, Technology, and Policy, NIST, input from our nonprofit advisors, as well as the major AI companies.

Our hope is to educate, to engage a wide audience, address vulnerabilities, and importantly, to grow a new profession. However, there is currently little to no investment in training and upskilling for this nascent practice. Our efforts rely heavily on volunteer effort, corporate or nonprofit funding.

Legal Protections for AI Ethical Hackers

[5] https://www.dfs.ny.gov/system/files/documents/2022/04/NYC_Council_Automated_Employment_Decision_Tools.pdf

[6] https://blog.twitter.com/engineering/en_us/topics/insights/2021/algorithmic-bias-bounty-challenge

[7] https://www.technologyreview.com/2022/10/20/1061977/ai-bias-bounty-help-catch-unfair-algorithms-faster/

Beyond financial and institutional investment, independent hackers need legal protection. Traditional ‘red teaming’ only happens under company purview. This is necessary but insufficient. “Ethical hackers” - that is, people who independently identify flaws in technology - are beneficial to society.

The history of information security[8] is riddled with stories of ethical hackers investigated and silenced for their service. Efforts like disclose.io highlight how federal anti-hacking laws are at odds with ethical hacking, and we envision the same problems arising for third party AI hackers.

In drafting new AI regulations, policymakers may well be tempted to create new restrictions on hacking AI. Policymakers should be very cautious about doing so. Hacking AI to cause harm is already illegal, and we must ensure that independent security research is protected. In the past, global efforts to ban all hacking have backfired.[9]

Recognizing this, the federal government has made significant progress in the past 3 years on providing researchers with protections under DMCA and CFAA, and in promoting adoption of CVD/VDP. The government should continue and expand on this policy[10]. More needs to be done in the following areas:

1) We need to ensure that these protections are explicitly extended to the context of AI. Any new AI regulations should allow for independent security research.

2) The DOJ's CFAA charging policy, though welcome, does not affect civil liability under the CFAA. One of the greatest risks researchers face are private lawsuits. CFAA reform in this area is needed.

3) The DMCA and CFAA improvements do not affect state laws. Yet state laws are often even more vague than CFAA. State law liability is another source of risk for security researchers.

In alignment with a robust, protected, and supported third party ecosystem, further development in both domestic and global government oversight is warranted to round out a well-designed governance ecosystem.

III. Creation of Domestic Body focused on Responsible Technology Use

A centralized body, an “Office of Responsible Use of Technology” could assist existing standards, oversight, and regulatory bodies by providing services and staffing to augment topical subject matter expertise.

A centralized body would serve many purposes: promoting interoperable licensing and oversight approaches, conducting empirical research into the effects of AI systems on American citizens to inform policy, sharing best practices across siloed government entities, and acting as a unified voice for the United States’ approach to responsible use of technology both domestically and internationally.

Parallels already exist in other governments. In the UK, groups like the Center for Data Ethics and Innovation - of which I am a board member - provide similar services. In the EU, the recently launched

[8] https://www.washingtonpost.com/news/powerpost/paloma/the-cybersecurity-202/2018/08/03/the-cybersecurity-202-the-law-doesn-t-protect-ethical-hackers-this-new-project-could-help-close-that-gap/5b6330421b326b0207955ecb/

[9] https://cdt.org/insights/draft-car-safety-bill-goes-in-the-wrong-direction/

[10] https://www.youtube.com/watch?v=duz7UXxR7tc

European Center for Algorithmic Transparency is tasked with similar responsibilities across all European Union nations.

IV. Support and Participation in Global Oversight Body

Finally - there is a sustained and increasing call for global governance of AI systems. Among them, I published an article in April[11] outlining an IAEA for AI, at his May Congressional testimony, OpenAI CEO Sam Altman called for similar governance[12], and recently, we have an op-ed by former NZ Prime Minister Jacinda Ardern[13] providing a high level blueprint for multistakeholder governance.

A global governance entity must respect national sovereignty while addressing the biggest risks introduced by AI systems. A diversity of governance approaches is generally better at identifying and addressing localized or context-specific problems. In identifying what is within the remit of a global entity, we must ask ourselves - "What is the climate change of AI?" - that is, what are the cross-border problems that are too big to be solved by a single company, a single country, or a single group?

My recommendations are as follows:

- First, a global governance body should only address the biggest cross-national problems
- Second, this body should consist of governments, civil society, and corporations
- Third, this group should be tasked with promoting benefit to society and human flourishing[14]
- Fourth, this group should design innovative solutions, not simply admire the problem
- Fifth, this group should have some degree of enforceability where appropriate

V. Summary

In sum - innovation in the national interest starts with good governance. By investing in and protecting this ecosystem, we will ensure that AI technologies are beneficial to all. Thank you for your time.

[11] https://www.wired.com/story/ai-desperately-needs-global-oversight/
[12] https://www.washingtonpost.com/technology/2023/05/16/sam-altman-open-ai-congress-hearing/
[13] https://www.washingtonpost.com/opinions/2023/06/09/jacinda-ardern-ai-new-zealand-planning/
[14] https://thehill.com/opinion/technology/4047323-artificial-intelligence-doesnt-have-to-be-inhumane/