Tech Global Institute
8 Brunswick Street
Brampton, ON L6X 4Y6

Arati Prabhakar, Ph.D.
Director, Office of Science and Technology Policy
The White House
1600 Pennsylvania Ave NW
Washington, DC 20500

July 6, 2023

**Comments from the Tech Global Institute on the OSTP National Priorities for Artificial Intelligence Request for Information**

Dr Prabhakar,

Thank you for requesting public comments on the important topic of setting U.S. national priorities and future actions on artificial intelligence (AI). Tech Global Institute is a global policy lab with a mission to reduce equity and accountability gaps between Big Tech/emerging technologies and the Global South (referred to as the Global Majority in this document). We are a community of senior policy and legal experts, trust and safety professionals, AI systems researchers and human rights specialists with experience working at leading U.S. technology companies, academia and multilateral organizations such as the United Nations. We focus on elevating the voice of underserved communities around the world in the design, development, deployment and governance of technologies that impact their lives.

With the emergence of new branches of large-scale consumer technologies, governments around the world, including the United States, are at a crucial crossroads to develop proportionate and effective guardrails that promote innovation while safeguarding democracy, privacy and civil liberties. Often these conversations do not adequately account for the *lived* experiences with technology among historically marginalized groups, especially those in low- and- middle-income countries, or the Global Majority. Specifically, technology infrastructure and product discussions about low- and- middle-income countries are limited to access and inclusion, although evidence on the use of social media and other products from the past decade has shown diverse applications, including in e-commerce, recruitment, healthcare, climate and agriculture. Subsequently, harms from these same technologies have disproportionately fallen on populations in low- and- middle-income countries, as evident through events in Myanmar, Philippines, Ethiopia, Sri Lanka and Bangladesh. Through updating its national priorities and future strategies on AI, the U.S. has the unique advantage of leading the world in the current and next phases of AI development, ensuring risks are mitigated early on and *all* parts of society can equitably reap its benefits.

In our submission, we address critical considerations in how AI impacts the lives of historically marginalized and underserved groups, the risk mitigation strategies and the role of the U.S. in working with international partners in developing responsible, equitable and inclusive AI systems. We welcome this opportunity to share evidence and look forward to future engagements on this critical topic.

**Specific Comments**

**1. What specific measures – such as standards, regulations, investments, and improved trust and safety practices – are needed to ensure that AI systems are designed, developed, and deployed in a manner that protects people's rights and safety? Which specific entities should develop and implement these measures?**

**AI ethics should be grounded in robust international human rights frameworks:** AI governance initiatives branded as "AI ethics" or "responsible AI" are based on the philosophical discipline of ethics. However, ethics is a malleable concept lacking universally agreed normative foundation, indivisibility, and enforceability inherent in human rights standards. Admittedly, international human rights laws were not designed to directly address concerns around AI and are insufficient to act as an entire system for the ethical management of AI. Despite these limitations, the underlying principles are a *universally* agreed blueprint for protecting human values and can be adapted to new contexts and evolving social norms. Human rights standards, such the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and the United Nation Guiding Principles for Business and Human Rights should therefore serve as the baseline for normative constraints on AI, supplemented by other ethical guardrails to create a comprehensive and complementary governance system.

The government's human rights commitments entail an obligation to actively engage in AI governance. This not only includes protecting individuals from rights abuse by commercial and non-state actors, but also its own compliance with human rights in AI-assisted systems that a government adopts. Critics often characterize human rights as roadblocks to innovation, citing them as being vague, static, and ineffective. We argue they are enablers of innovations through the promotion of responsible practices and rights-respecting standards. International human rights laws are grounded in a well-calibrated approach that considers the legality, necessity, and proportionality of rights limitations, as well as an ecosystem for provision of redress for the public in case of violations, thereby offer a predictable, progressive, and potent benchmark for AI governance. It can address concerns around privacy, data protection, equality, non-discrimination, and socio-economic rights.

As underscored by the Guiding Principles on Business and Human Rights that was unanimously endorsed by the UN Human Rights Council and General Assembly in 2011, corporations and other non-government entities have a responsibility to respect human rights. Companies should have a policy commitment to meet their human rights responsibilities, approved at the most senior level, publicly available, and embedded in the culture of the business. Furthermore, companies must also have an ongoing due diligence process of human rights impact assessment, tracked for responsiveness and reported externally, which allows them to identify, mitigate and remedy human rights impacts.

**AI systems should be categorized and assessed on the basis of their impact:** Adoption of a well-defined risk-based regulatory approach will ensure a predictable and proportionate response to risks while avoiding unnecessary restrictions on innovation and investment. Depending on the level of risk – to be assessed by the impact of a particular AI system on individuals, with special attention paid to historically marginalized groups – regulations should either be in the form of complete prohibitions or constraints on applications of AI.

An impact-based risk assessment for AI applications involves evaluating the potential *effects*, probability of (re-)occurrence, and severity of risks associated with the system, as opposed to the *purpose* of the AI application. This approach provides a comprehensive and actionable understanding of the potential harms that may arise, as well as the necessary measures to mitigate and manage the identified risks, thereby ensuring a general purpose AI system undergoes the same rigorous assessment as a specialized system. The evaluation process includes hazard identification, vulnerability assessment, impact analysis, risk evaluation, and risk management. An AI system that could adversely impact individuals should be classified accordingly, and corresponding regulatory measures should be imposed.

For instance, the EU's AI Act (AIA) prohibits certain manipulative, exploitive, and invasive AI systems due to their serious ethical and human rights impacts. Certain "high risk" applications must undergo conformity assessments, periodical audits, and evaluation throughout the systems' life cycle, in addition to complying with data governance, documentation and record keeping, transparency, human oversight, and security requirements. For the low-impact AI systems, there are basic transparency, labeling, and voluntary compliance obligations.

Governance norms should carefully and thoughtfully place the needle in the spectrum between pro-innovation and protection from potential harms. Because AI and its capability and shortcomings are inherently different to other technologies and significantly amplifies risks of harm, co- and self-regulatory approaches may not be adequate. We recommend enactment of a legally binding and harmonized framework that is future-proof, technology-neutral, innovation-friendly, and fit-for-purpose.

**AI systems should mitigate risks throughout the product life cycle:** Developers of AI systems should undertake ongoing due diligence for high-risk AI applications throughout the product life cycle, aimed at mitigating and managing potential risks and hazards associated with the system. In addition to the AI system itself, there should be assessments of risks associated with personnel involved and the specific processes behind developing the systems. Given the self-learning and self-improvement capabilities of AI, comprehensive impact assessments and ongoing audits are necessary, as some risks may only become apparent after the system is operationalized. Even if the algorithm is safe and ethical by design, the context AI use also plays a role in determining its impact, necessitating retrospective review of the system. Moreover, iterative development and improvement cycles that AI systems undergo requires integration of risk mitigation practices throughout the life cycle. Risks related to AI systems can also evolve over time due to changing circumstances, technological advancements, or new vulnerabilities, thus requiring continuous monitoring and review.

Decision-making structures should also be established to monitor, identify, and address human rights risks and ethical concerns associated with AI. Companies should fulfill their common law duty of care by identifying and mitigating actual and potential negative effects of AI systems. In particular, *human rights by design* should be applied to AI by systematically embedding human rights at every stage of the design, development, and deployment.

**AI oversight needs both a national and international approach:** At the national level, a dedicated federal agency should be established for AI supervision and regulation, to ensure accountability, compliance, and alignment of the AI technologies operate with the public interest. Currently, the regulation of digital platforms falls on multiple agencies, such as the Department of Justice and the Federal Trade Commission, which often lack *all* the necessary expertise and resources. To address this, following the precedent set by expert regulatory bodies like the Food and Drug Administration, the Federal Communications Commission, and the Consumer Financial Protection Bureau, we recommend the establishment of an expert regulatory agency specifically focused on AI. This agency should be empowered with investigative, enforcement, and regulatory powers, and extraterritorial mandate for cross-border oversight.

Furthermore, AI systems are often deployed globally, crossing national boundaries and affecting rights and values on an international level. Inconsistent regulatory approaches across jurisdictions can lead to confusion, legal uncertainties, hinder the development and interoperability, and pose risks to individuals' fundamental rights. Given the influential role and global impact of the U.S. in technology regulation and its global impact, it is essential for the government to promote harmonization of AI standards. This approach will facilitate smoother cross-border operations, foster global innovation, and ensure fairness, transparency, and accountability on a worldwide scale. Embracing an international and inclusive approach to AI regulation allows countries to collaboratively address global challenges and harness technology for the benefit of all humanity, reflecting diverse perspectives and a more comprehensive understanding of societal values and ethical principles. This approach would also avoid duplicating efforts and promote the accelerated development of policies that keep pace with AI advancements.

**Transparency is critical to achieving effective and improved AI accountability:** Transparency enables accountability and can yield a positive effect on understanding and addressing the societal impacts of AI systems. Companies are in a unique position to inform the public about social harms given certain safety-related use cases, given the company's unprecedented tools and data. For example, if the OpenAI API is being used widely for child-related purposes, this can improve public safety. According to Mozilla's AI Transparency in Practice, transparency improves accuracy and target goal achievement, facilitates new insights by investigating learned prediction strategies, helps avoid unwanted outcomes, avoids bias, justifies decision-making to users and other stakeholders, enables user control, increases security, verifies generalizability of the models, and improves system robustness.

Upcoming legal frameworks, notably the AIA, the Digital Services Act, the UK Online Safety Bill, potentially the U.S. Platform Accountability and Transparency Act, and others, include provisions around transparency and the documentation of AI systems and will increasingly require transparency obligations for AI services and products. Specifically, reporting on areas such as access to data, algorithmic transparency, in-product features, process descriptions, public risk mitigation plans and content moderation will be critical to build trust and improve oversight of AI systems.

The most critical information that companies should share externally, listed in the order that we would recommend them, includes explanation surfaces or dashboards around:

*Data governance (data privacy and intellectual property issues)*: Algorithmic transparency can include understanding why or why not an algorithmic system recommends a particular outcome, knowing when and when not to trust an outcome, when it succeeds or fails, how to opt out, override, or correct the AI system's output, and insights into model behavior. The draft AIA requires transparency around data sources, data governance, copyrighted data, compute, energy, capabilities and limitations, risks and mitigations, evaluations, testing, machine-generated content, member states, and downstream documentation.

*Bias and individual harms mitigation*: Transparency around harm mitigations involves making users aware of system limitations and model reporting (i.e. system cards). OpenAI's recently published GPT-4 System Card describes the safety challenges that arise from GPT-4 and explains the interventions implemented to mitigate potential harms from its deployment, namely hallucinations, harmful content, harms of representation, allocation, and quality of service, disinformation and influence operations, proliferation of conventional and unconventional weapons, privacy, cybersecurity, potential for risky emergent behaviors, interactions with other systems, economic impacts, acceleration, and overreliance.

*Societal concerns (environment, job displacement, trustworthiness, etc.)*: The potential impact on transparency here is extensive and well documented, given its direct connection to safety concerns expressed by various stakeholders.

*Alignment and dangerous capability monitoring*: This reporting can cover auditing for capability monitoring and evaluating the training set-up to see if the model can self-replicate. These elements correspond directly to commonly expressed short- and long-term concerns from the AI safety community.

**4. What are the national security benefits associated with AI? What can be done to maximize those benefits?**

**AI systems can improve and automate intelligence gathering and analysis:** AI can be used to process large amounts of data and identify patterns that would be difficult or impossible for humans to detect. Natural language processing and computer vision techniques can extract relevant information, identify patterns, and generate insights to support intelligence agencies in their mission to assess threats and plan operations. This can be used to improve intelligence gathering and analysis, which can help to prevent and respond to threats, and be transformative in accelerating U.S. military and information superiority. One specific benefit is the use of AI to automate tasks that are currently performed by humans, such as monitoring systems and responding to threats. This can free up human resources for other tasks, such as strategic planning, complex analyses and decision-making.

**AI systems can enhance cybersecurity and enhance situational awareness:** AI can be used to bolster cybersecurity efforts by detecting and responding to cyber threats more effectively. Machine learning algorithms can analyze network traffic patterns, identify anomalies, and detect potential cyberattacks. AI can also assist in developing proactive defense mechanisms and improving incident response. This can help to protect critical infrastructure and data from malicious actors. AI can process vast amounts of data from various sources, such as sensors, satellites, and social media, to provide real-time and comprehensive situational awareness. This enables better threat detection, early warning systems, and decision-making support for defense and intelligence agencies.

**AI can enable and accelerate development and deployment of autonomous systems:** AI enables the development of autonomous systems, such as unmanned aerial vehicles (UAVs) or unmanned underwater vehicles (UUVs). These systems can perform tasks like surveillance, reconnaissance, and logistics without putting human operators in harm's way, thereby increasing operational efficiency and reducing risks.

**AI can bolster predictive modeling and risk assessment, thereby supporting more efficient decision-making:** AI algorithms can be used to create predictive models and conduct risk assessments for various scenarios, including geopolitical events, terrorist activities, or natural disasters. This can help national security agencies allocate resources effectively, plan responses, and mitigate potential risks in advance. AI can be used to provide decision-makers with better information and insights, which can help them to make more informed decisions. This can be particularly important in complex and rapidly changing situations.

To maximize these benefits, the United States can consider the following actionable steps.

**Invest in AI research and development:** Governments should prioritize funding for AI research and development in the defense and intelligence sectors. This will enable the exploration of new AI applications, the development of cutting-edge technologies, and the training of skilled personnel.

**Build collaboration between government and industry via public-private partnerships:** Governments should foster partnerships and collaborations with the private sector to leverage AI expertise and access advanced technologies. This can be done through initiatives like public-private partnerships, joint research projects, and technology transfer programs.

**Develop ethical and responsible AI:** National security agencies should ensure that AI development adheres to ethical guidelines and principles. This includes transparency, fairness, and accountability in AI algorithms and decision-making processes to prevent unintended biases or risks.

**Bolster data sharing and integration:** Effective utilization of AI in national security requires access to diverse and high-quality data. Governments should promote data sharing and integration across agencies to enhance the accuracy and reliability of AI models and systems.

**Improve robust cybersecurity measures:** As AI becomes increasingly integrated into national security infrastructure, robust cybersecurity measures must be implemented to protect AI systems from potential attacks or manipulation. This includes secure data storage, encryption, and continuous monitoring of AI systems.

**Develop a skilled workforce:** The public needs to be educated about AI and its potential benefits and risks for national security. This will help to build public support for AI-enabled technologies and ensure that they are used in a responsible manner. Governments should invest in training programs to develop a skilled workforce proficient in AI technologies and applications. This includes attracting and retaining AI experts, promoting STEM education, and fostering collaboration between academia, industry, and government.

**Partner with allies:** The United States should partner with allies to share information and resources on AI for national security. This will help to ensure that AI is used to protect the common good.

**Monitor the development of AI:** The government should monitor the development of AI and its potential implications for national security. This will help to identify potential threats and vulnerabilities and develop strategies to mitigate them.

**7. What are the national security risks associated with AI? What can be done to mitigate these risks?**

**AI systems are vulnerable to adversarial attacks:** AI systems can be vulnerable to adversarial attacks, where malicious actors manipulate or deceive AI algorithms to produce incorrect or harmful outputs. This can impact critical systems like autonomous vehicles, surveillance systems, or decision-making algorithms. AI could be used to launch cyberattacks on critical infrastructure and data systems. This could disrupt essential services and cause economic damage.

**AI systems pose unprecedented privacy breach risks:** AI relies on large amounts of data, including sensitive and classified information. The security and privacy of this data become paramount concerns, as unauthorized access or data breaches can compromise national security, intelligence sources, or military operations.

**AI can exacerbate bias and discrimination that disproportionately impacts certain groups:** AI algorithms can inherit biases present in training data, resulting in discriminatory outcomes. Biased decisions in areas such as law enforcement, immigration, or counterterrorism could disproportionately impact certain groups and undermine trust in national security institutions. Synthetic media refers to manipulated or fabricated content, including images, videos, or audio, that convincingly imitates real individuals or events. The sophistication of AI algorithms makes it increasingly challenging to distinguish between real and synthetic media, raising concerns about its potential discrimination.

**Interconnected AI systems are more prone to systemic vulnerabilities:** Increasing reliance on interconnected AI systems creates vulnerabilities that adversaries can exploit. Attacks targeting critical infrastructure, communication networks, or command and control systems could disrupt operations, causing significant national security implications.

**Autonomous weapons systems can fuel an arms race:** AI advancements may lead to the development of autonomous weapons systems, which can potentially change the dynamics of warfare. The uncontrolled proliferation of such systems may fuel an arms race and raise concerns about the ethical and legal implications of their use.

**Synthetic media can be used to spread disinformation that would erode trust in institutions:** AI could be used to spread disinformation and propaganda. This could undermine public trust in government and institutions, and could lead to social unrest. In the context of democratic policies, such as elections, this poses a significant challenge as it can erode trust in the authenticity of information and undermine the integrity of the electoral process. Malicious actors can exploit deepfakes to fabricate compromising or damaging content about political candidates, spreading misinformation and sowing discord among voters. Deepfakes can be used to generate false narratives, mislead the public, and manipulate public opinion, ultimately compromising the democratic decision-making process. The potential consequences include widespread distrust, polarization, and a loss of faith in democratic institutions. Foreign interference, especially during elections, is more aggregated with AI, particularly in high-priority / more sensitive foreign markets (India, Taiwan, China, etc.).

**AI can assist evasion of detection of child pornography:** One of the national security risks associated with AI in the context of Child Sexual Abuse Material (CSAM) violation is the potential for AI to be used to evade detection and distribute CSAM. AI-powered techniques can be employed to manipulate or generate new content that resembles CSAM, making it challenging for traditional content moderation systems to identify and remove such material. The advancement of AI algorithms could enable offenders to automate the production and distribution of CSAM at a larger scale and with greater efficiency, thereby exacerbating the problem. Additionally, AI may be utilized to anonymize or obfuscate CSAM, making it harder to trace the origin and identify those involved in its creation or dissemination. Furthermore, the use of AI-generated deepfake technology poses a serious threat, as it can be employed to create highly realistic and deceptive CSAM, potentially leading to false accusations or undermining the trust in authentic evidence.

Risk mitigation strategies include:

**Bolster robust security measures:** Implement strong security protocols, encryption techniques, and authentication mechanisms to protect AI systems and data from unauthorized access or manipulation. Continuously monitor and update security measures to address emerging threats.

**Invest in AI security research:** The government and private sector must invest in research on AI security. This research should focus on developing ways to protect AI systems from cyberattacks and other malicious threats. Investing in AI security research goes beyond just technical expertise. It is crucial to involve non-engineers, such as policy experts and safety professionals, to address the broader implications of AI. Collaborative efforts between multidisciplinary teams can ensure a holistic approach, incorporating ethical considerations, policy frameworks, safety measures, and responsible deployment guidelines to safeguard against potential AI risks.

**Build adversarial defense:** Develop techniques to detect and defend against adversarial attacks. This includes methods such as robust training, anomaly detection, and incorporating adversarial resilience into AI algorithms to make them more resistant to manipulation.

**Promote ethical and transparent AI:** Promote the development and deployment of AI systems that adhere to ethical guidelines, transparency, and accountability. Ensure that AI systems are designed to minimize biases, undergo regular audits, and provide clear explanations for their decisions.

**Bolster data governance and privacy:** Establish stringent data governance frameworks to protect sensitive data, ensure privacy compliance, and define clear guidelines for data collection, storage, and sharing. Implement mechanisms for obtaining informed consent and anonymizing data where necessary.

**Foster international collaboration and norms:** Foster international cooperation to establish norms and regulations governing the development and use of AI in national security. Encourage dialogue between nations to address concerns, define standards, and prevent an uncontrolled arms race in autonomous weapons.

**Maintain human oversight and control:** Maintain human control and decision-making in critical national security operations. Ensure that AI systems are designed to augment human capabilities, allowing humans to intervene, verify, and override automated decisions when necessary.

**Promote education and awareness:** Promote public and professional education on AI's risks and benefits in national security. Develop training programs to enhance the understanding of AI technologies, ethics, and security among defense personnel, policymakers, and the general public.

**10. What are the unique considerations for understanding the impacts of AI systems on underserved communities and particular groups, such as minors and people with disabilities? Are there additional considerations and safeguards that are important for preventing barriers to using these systems and protecting the rights and safety of these groups?**

**AI systems should bring rights at the center of design:** While AI and AI-empowered systems and technologies can help particular groups in various ways, they need to be carefully applied since these systems are prone to have bias and they reproduce systemic inequalities in social fabrics. Inclusion, safety, transparency, and fairness should be first principles while designing and deploying such technologies. There are existing frameworks for the protection of children and youth, and disabled persons, that already have universal legitimacy including The UN Convention on the Rights of the Child (CRC) treaty and Convention On The Rights Of Persons With Disabilities (CRPD) treaty. The U.S. needs to consider these frameworks as baselines to establish rights of minors and people with disabilities in relation to AI development and deployment. At the same time, the U.S. government should collaborate with global bodies such as UNICEF, Amnesty International and OHCHR to better understand how AI systems can protect, provide for, and empower children, youth and persons with disability particularly those who are in the Global Majority. However, building minimum universal standards for ethical AI for these groups might not be enough, they need to be ratified strongly to block high-risk AI applications that pose possible risks of violating the rights of vulnerable groups.

**AI systems should incorporate *safety by design* and *privacy by design* to reduce likely harms to minors, including production of child sexual abuse materials:** Rapid development of AI systems without adequate guardrails on data collection and labeling for training data pose disproportionate risks for minors, especially in terms of their privacy, exposure to harmful or age-inappropriate content, bias and discrimination and exploiting their vulnerabilities to target of unethical or manipulative advertising. Specifically, diffusion models, such as Stable Diffusion and OpenAI's Dall-E, are being used to produce child sexual abuse materials (CSAM). An emerging concern in the field of child safety is not only the creation of illegal synthetic media of real children but also to make sexual abuse materials of children that do not exist. This phenomenon will likely inundate databases like National Center for Missing and Exploited Children (NCMEC) that will further complicate efforts to differentiate real victims. Although the U.S. has existing standards that bans child pornography irrespective of how the image was created, which includes AI systems, it is yet to be tested in court at the scale at which generative AI systems can and are producing CSAM.

It is therefore essential that AI ecosystems, including generative AI, incorporate *safety and privacy by design* principles throughout the product lifecycle to ensure thoughtful development and deployment. Specifically, systems should remove harmful content for minors from training data, for example through hashing and matching techniques, as well as detect harmful input and output prompts. Additionally, AI systems should adopt robust data protection policies, which includes measures ensuring strong encryption and anonymization techniques to protect individuals' privacy, and enable individuals to access, rectify, and erase personal information, especially if it concerns a minor. To facilitate the exercise of these rights, AI systems should have user-friendly interfaces that allow users to easily navigate privacy settings and manage their personal information.

**Protecting the persons with disability from negative impacts of AI:** AI systems may discriminate against facial differences, gestures, speech impairment and other forms of disability. Since recruiting processes increasingly use algorithms to screen out candidates, the immediate effect of these biases and exclusionary processes will be, in many situations, to impede the job rights of disabled persons. Many countries have already started making important decisions regarding who receives public services like social safety benefits or health insurance using predictive AI models, which in many cases are exacerbating biases and can disproportionately affect people with disabilities. Using AI for biometrics could put disabled individuals in dangerous positions, especially when it decides whether or not people should move across nations or regions during emergencies and humanitarian crises.

In order to protect the rights of disabled people, we recommend the U.S. government to give the Convention on the Rights of Persons with Disabilities (CRPD) treaty priority when developing and implementing AI tools that directly affect those at-risk groups, such as biometrics, facial recognition technology (FRT), and emotional recognition technology (ERT). It should work with the OHCHR and other international human rights organizations to adopt more accessible rules and guidelines for the businesses who create AI tools and to compel those businesses to make the reasonable adjustments that people with disabilities require.

**Ensuring skills as well as data and ethics literacy:** To make AI and data driven systems systems inclusive, the Global Majority population need to be resource-ready. There should be mandates and global priority to reduce inequality in terms of access, skills, and outcomes and ensure equitable representation of the world as data and control over data flows by people to achieve best usages of Data for Development (D4D) or AI for Development (AI4D) narratives. While national data centers are often cited as a data sovereignty measure, the evidence indicates the location of storage does not have significant bearing on people's access or rights to their data. Instead, it risks fragmenting the open, harmonized internet which likely adversely affects resource-constrained populations in the Majority World. It is therefore critical to invest in human capital, not only in terms of practical skills to understand and analyze data to better evaluate AI systems, but also in data literacy that empowers people to think more critically about their data rights. Ethics fundamentals including transparency and explainability should be mainstreamed to ensure diverse populations can respond and interact with AI systems in a more informed, responsible way.

**Making AI models and systems socially and culturally situated and responsive:** AI systems including LLMs need to be sensitive to multilingualism and cultural diversity existing in the Global Majority. Common Crawler (used for most AI systems) uses only textual languages on the internet, which is predominantly in English. This means that most content is in one language and in a limited format. These limitations on how the data is collected need to further assessed to ensure AI systems are not skewed to one part of the world and ineffective, or even harmful, for the rest. Language models developed through other modalities (image, voice-based, etc.) are more appropriate for the Global Majority audience.

At the same time, regulations should be in place to safeguard vulnerable groups from bias and discrimination, especially those impacting class, gender, sexual identity, nationality, religion and ethnicities. AI models such as LLMs can rapidly learn demeaning language and harmful stereotypes about groups who are frequently marginalized. Training data often reflects historical patterns of systemic injustice, which can be compounded for certain intersectionalities, around gender and race. AI development should therefore be evaluated not only on the fairness of the technical systems, but also the environments in which it is developed, the diversity of the AI development team and their responsive to languages and multi-linguality.

**11. How can the United States work with international partners, including low- and middle-income countries, to ensure that AI advances democratic values and to ensure that potential harms from AI do not disproportionately fall on global populations that have been historically underserved?**

**Responsible AI needs to be defined for the Global Majority:** Responsible AI, as a principle, needs to be integrated as a core value and practice while designing AI systems, not as a set of high level principles that do not translate to practical measures. That said, it is critical to acknowledge that existing and emerging policy and procedural guardrails on responsible AI do not exacerbate or ignore the power dynamics that operates through AI systems and that a one-size-fits-all approach to AI fairness cannot achieve desired results in resource-constrained, underserved societies, particularly in low-income, immigrant communities in the U.S. and low-and- middle-income countries. Responsible AI needs to be defined from ethical and legal point of view specifically in terms of what it means in the context of the Global Majority At present, it is largely designed and used in a narrowly scoped, Western context exclusively applicable to and responding to urban, white and wealthier demographics.

We therefore recommend: first, AI systems need to be more carefully designed, developed and deployed in the Global Majority contexts with efficacy, robustness or reliability to avoid unfair treatment of individuals or groups who are already marginalized. Second, AI systems should provide and communicate decisions, terms of service or predictions that *diverse* local users, skilled and semi-skilled developers, and local and national regulators can understand, taking into account the level of digital literacy and social reality of post-colonial, resource-constrained regions. Third, privacy-enhancing techniques to mitigate personal or critical data leakage should also be developed with required Global Majority contextualisation. Finally, applications of AI technologies might require low-tech innovations which are typically undervalued and underinvested by Global North-centric companies. Not implementing a project which is high-tech in nature but brings limited benefits for the local communities can also be a responsible step, particularly in the Global Majority contexts.

**The U.S. should work with international partners to mandate impact and fairness assessments of AI systems:** Before implementing AI-driven projects at scale, particularly outside of the U.S., it should be mandatory for companies to undertake impact assessment processes and publicly share results, including inviting comments from stakeholders. This is important because AI systems can dominate, exploit, and even dehumanize vulnerable populations in such contexts in an imbalanced way and introducing repairs and post-hoc fixes can often be ineffective to resist those harms. These assessments should prioritize the issues that are critical for underserved societies, specifically among ethnic, sexual and religious minorities, as well as low- and- middle-income countries. Attention should be paid to topics on digital sovereignty, infrastructural and regulatory capacities, harms associated with the labor and material supply chains of AI technologies, beta-testing of new features and products, and exploitative commercial use.

Recent investigative reports reveal how the global AI industry is exploiting nefarious strategies to violate such citizen rights for their profit making. South Africa's foreign corporations use AI technologies to monitor its people, exporting surveillance data for racial control. Venezuela's AI industry seeks cheap labor during economic crises, while Jakarta's Gojek taxi drivers fight back by strengthening worker power through community actions. The U.S. in collaboration with its international partners should put increasing pressure on corporations to achieve required ethical standards and public impact assessment reports will play a critical and concrete role in this regard.

**The U.S. should work with international partners to establish labor protections:** The U.S. Government along with its international partners, including the World Bank, United Nations and the European Union should develop foundational standards for strengthening labor protections for low-skilled data labeling and content moderation workers. These workers are predominantly invisible and situated in low- and middle-income countries in Asia, Africa and South America, facing growing demands to manually tag videos, sort photos, transcribe video and label large troves of data that is the bedrock of any AI system, including e-commerce, chatbots, voice assistants and self-driving cars. Researchers and experts have dubbed this as ghost work whose market value is estimated to reach $13.7 billion by 2030. The AI crowdworking platforms are largely unregulated with no existing safeguards for their contractors or gig workers directly employed by AI companies on minimum pay, parental leave or working conditions, sometimes earning as low as 11 cents per hour. Addressing these systemic gaps should be a priority since the extractive nature of AI has already been reported, discussed and criticized by the academia and the media.

**The U.S. should work with the United Nations, International Organization for Standardization and the World Trade Organization, on developing international rules to govern AI:** By working together with partners, the U.S. can contribute to the development of consistent international standards and norms for AI. This reduces fragmentation and conflicting regulations across different jurisdictions, facilitating smoother international trade and cooperation. Given the transnational nature of AI systems, it is imperative to ensure the process of developing such frameworks is diverse and inclusive, factoring in the unique societal, privacy, cybersecurity and accountability challenges of different regions. Specifically, by engaging with the WTO, the U.S. can contribute to the establishment of rules and regulations that reduces barriers to trade, protects intellectual property rights, addresses privacy concerns and facilitates market access for compliant AI systems and services. As part of this process, it is critical to engage with diverse civil society and academic institutions in Global Majority contexts to ensure transparency, accountability, diversity of perspectives and ownership in the formulation and implementation of any international AI frameworks or acts.

**The U.S. should work with low- and- middle-income countries to develop guardrails on AI technology transfers:** In any technology-transfer partnership, the questions of geopolitical power imbalance should be brought into the forefront since AI development and applications will be a deeply divisive if Global Majority countries cannot access the critical digital infrastructures and resources required for participating in AI innovation. Countries need to agree on the existing structural inequality in developing governance and legal frameworks for AI and should find ways of mutual cooperation to reduce them. Actors and decision-makers from the Global Majority must have control over the strategic decisions of application of AI technologies in their own contexts so that AI practices cannot cause more harm than benefits for those stakeholders, including partnering with the U.S. on research funds and resources to address both innovation and ethical implications. The U.S. should play a leading role in identifying mechanisms and protocols that ensure meaningful participation of underserved stakeholders and reduce institutional barriers for such participation. At the same time, the Global Majority stakeholders need to engage in formally defined roles to collectively participate in the global AI technology transfer and governance processes.

**The U.S. should work with international partners to invest in AI infrastructure more broadly:** Standardization and harmonization of rules on AI infrastructure will be critical to achieve a shared understanding of AI systems to safeguard the open internet. The U.S. should engage with international efforts, as well as International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) to ensure interoperability, ethical standards and safety across borders. A common understanding will bolster globally distributed AI infrastructures to be more resilient to malicious or adversarial actors and nation states. In addition to developing a shared vocabulary, the U.S. should establish partnerships with allies and low- and- middle-income countries to foster AI research and innovation, jointly fund AI infrastructure projects, enable AI talent mobility across borders through academic exchanges and industry secondments, and pool financial resources to strengthen computing resources.

**21. What are the global labor force implications of AI across economies, and what role can the United States play in ensuring workforce stability in other nations, including low-and middle-income countries?**

**Widespread AI adoption will have a disproportionate impact on outsourcing jobs:** The widespread adoption of AI technologies is expected to automate various routine and low-skilled tasks across industries, especially those that rely heavily on outsourcing. Sectors such as manufacturing, transportation, customer service, and other outsourcing-dependent industries may experience significant job displacement due to the automation of predictable physical or cognitive tasks. As AI streamlines these tasks, the need for human involvement decreases, potentially leading to workforce reductions in outsourcing destinations. This can have significant implications for low- and- middle-income economies that heavily depend on outsourcing as a source of employment and economic growth. Displaced workers in these sectors will need to acquire new skills or transition to different job roles to remain competitive in the changing job market.

While AI automation may eliminate certain tasks, it can also transform job roles by augmenting human capabilities. AI technologies can enhance productivity and decision-making, leading to the creation of new jobs that require complementary skills. The demand for skills such as data analysis, machine learning expertise, and AI programming is likely to increase, along with the need for specializing in AI ethics, policy, regulation, critical thinking, creativity, and emotional intelligence. Additionally, AI can facilitate the development of innovative outsourcing solutions, enabling companies to offer more specialized and value-added services to their clients. The U.S. can establish partnerships with other countries to focus on workforce upskilling and reskilling initiatives, including enabling academic exchanges, establishing industry secondments and investing in upskilling programs.

**Widespread AI adoption can exacerbate income inequality that would require policy intervention such as tax-benefit systems:** The impact of AI on the labor force can contribute to income inequality and job polarization. High-skill occupations that leverage AI technologies may experience wage growth, while low-skill jobs face wage stagnation or decline. The digital divide and access to AI tools and education may exacerbate these disparities, creating challenges in ensuring equitable opportunities and distribution of benefits across society. The U.S. can consider creating supportive environments for workers through effective social safety nets, lifelong learning opportunities, and policies that address income redistribution, labor rights, and the ethical use of AI. By collaborating with international partners, the U.S. can explore tax-benefit systems to help workers in low- and middle-income countries with transitions to new opportunities in different occupations and sectors.

**Labor protections are critical for AI workers in low- and- middle-income countries:** With rapid advances in AI systems, there is a significant spike in demand for data labeling and annotation jobs that are predominantly situated in low- and- middle-income countries. At present, crowdworking is an unregulated industry that does not have any worker protection measures. The U.S. can play a crucial role in supporting other nations by advocating for labor protection laws in areas such as enacting and enforcing worker protection regulations, ensuring occupational health and safety standards, and establishing fair minimum wage policies. By collaborating with international partners, the United States can actively contribute to ensuring that workers in low- and middle-income countries are not only provided with job opportunities but also enjoy essential labor rights. These collaborative efforts can help foster fair wages, safe working conditions, and robust social protections, promoting the stability and well-being of workers globally.

**Foreign assistance should prioritize workforce stability and entrepreneurship development:** Through international development and aid programs, the U.S. can contribute to workforce stability in low- and middle-income countries. This can involve funding initiatives that focus on skills training, job creation, and entrepreneurship development, with a particular emphasis on sectors with growth potential and alignment with local needs and priorities**.** This can involve providing mentorship, access to capital, and business development resources to individuals and start-ups in low- and middle-income countries, enabling them to establish sustainable enterprises and contribute to local job creation.

Additionally, through collaborations with international organizations, such as International Labor Organization, regional bodies, and other countries, the U.S. can share best practices, influence policies, and coordinate efforts to support sustainable economic development and workforce stability.

**Additional comments**

In addition to our specific responses above, we would like to draw attention to the field of AI ethics, which currently is predominantly developed and implemented in the U.S. and is missing critical views of the Global Majority. Existing AI ethics frameworks skew towards a technical solution in the *process* of designing and developing AI systems, such as performance, efficiency, novelty, data protection practices and quantitative evidence. Further, they tend to highlight societal value systems such as fairness, explainability, safety and transparency from a principled standpoint that has shortcomings in enforcement. AI ethics lacks a disciplined reinforcement mechanism, is not sensitive to different contexts and cultures, and is narrowly scoped. It does not account for challenges in multi-linguality, multi-mode data sources and multi-cultural nuances and therefore, cannot be universally applied. It also does not account for the environment on which AI systems are developed, which includes infrastructure gaps, labor protections, ecological footprint and global burdens of harms.

Future actions on AI should therefore take a more comprehensive view on the field of AI ethics that evaluates the AI system, as well as the diversity of the AI development team, the environment in which a system was developed, its human rights and labor implications, its effect on a historically underserved global audience, and its burden on climate change.

We are grateful for the opportunity to provide feedback.

Sincerely,

Sabhanaz Rashid Diya
Founder

Shahzeb Mahmood
Head of Legal

Sheikh Waheed Baksh
Head of Trust & Safety Programs

Theodora Skeadas
Senior Policy Advisor

Abdullah Hasan Safir
Leverhulme Centre for the Future of Intelligence
University of Cambridge

*On behalf of Tech Global Institute*