Inputs for the Discussion towards AI Safety Institute's Stakeholder Consultation

Digital Empowerment Foundation

To:
Mr. Abhishek Singh
Additional Secretary and CEO, IndiaAI Mission
Ministry of Electronics and Information Technology
Government of India

Date: 04 October,2024

Subject: Submission of points for the AI Safety Institute stakeholder Consultation

Dear Sir,

The Digital Empowerment Foundation (DEF) wishes to thank the IndiaAI mission, Ministry of Electronics and Information Technology, for the opportunity to submit our points for the AI Safety Institute Stakeholder Consultation. Digital Empowerment Foundation is a New Delhi-based not-for-profit organisation. It was born from the deep understanding that marginalised communities living in socio-economic marginalisation and information poverty can be empowered to improve their lives by providing access to information and knowledge using digital tools. In the present times, with the rise of AI technology in the digital ecosystem, DEF has a greater responsibility to ensure safe and trusted AI is recognised. We are grateful that the IndiaAI mission has sought stakeholder consultation on AI safety Institute under the IndiaAI mission. DEF would be happy to provide any further support to IndiaAI's mission on this issue.

Yours sincerely,

Osama Manzar
Founder and Director
Digital Empowerment Foundation
House 44, 3rd Floor, Kalu Sarai, New Delhi - 110017
Website: www.defindia.org

GENERAL INPUTS

The Government of India has announced its IndiaAI mission with a cabinet approval and an overall allocation of Rs. 10,371.92 Crores and a budget allocation of Rs. 551.75 Crores for the FY 2024-25. The major objectives outlined in the IndiaAI mission are ensuring AI safety and supporting innovative AI through collaborative engagements between public and private actors. While this mission focuses on establishing supercomputing environments for generative AI made by and made for India, the lack of background infrastructure and information for developing the LLMs and LMMs should be discussed in detail. There are no set guidelines or support systems for the early detection or counteraction of AI malpractices.

This gap could be addressed by the AI Safety Institute, which could work in tune with similar institutes formed in other nations. AI safety is not a domestic problem, and it cannot be addressed only at a national level. However, India might face issues arising out of the domestic characteristics of the nation and its society in the future if we stress too much on the foundational principles of AI developed by companies outside the country. Addressing this imminent danger could not be avoided only by setting high data safety and security standards, homegrown AI systems, and domestic HPCs (High-Performance Computing).

We suggest developing a comprehensive framework for AI-based systems under the AI Safety Institute with objectives that govern the decision-making skills of the AI systems developed in India, focusing more on the defects and limitations of the LMMs/LLMs being used. Much stress is needed not on the safety of AI systems, which is inherent in all digital systems, but on the possibility of bad decision-making by the AI systems due to the unknown/unavailable/limited training models. A strategic intervention is required at the policy level to avoid such lousy decision-making in any AI system through continuous training, ensuring end-user awareness of the limits, etc. The AI Safety Institute may also include this in its ambit.

SPECIFIC INPUTS

1. What should be the AI Safety Institute's focus?

a) What should be the core objectives of the AI Safety Institute?

India's AI Safety Institute should prioritise Data Quality & Management as well as Human Values and AI Ethics for formulating its objectives along with the innate goals of Governance and Oversight, Risk Management, and Education and Awareness. The latter three are still covered under the objectives of other existing institutions and organisations, such as the CERTin (Computer Emergency Response). AI Safety Institute shall function to build the gaps at the technical and implementational level of AI systems through global public-private collaborations.

An AI safety institute should be able to map out the different spectrum of issues both on the production side and consumer side of AI technology with a focus on developing possible institutions of grievance redressal. For example, on the consumer end, this can include issues of bias, malfunctioning, invasion of privacy etc. On the production side, it can include issues of excessive natural resource consumption, environmental impact, unethical data mining, etc.

What organisational structure will best support its mission and scalability?

The AI Safety Institute may function as an Inter-governmental organisation with the Central and State governments and their diverse functionaries related to the field as invitees. It should constitute a board of members with subject experts and other stakeholders for regular governance and action.

An organisational structure can be imagined as a multi-stakeholder body where there are state actors, private entities, policymakers, experts but also representatives from communities that are likely to be impacted by the harmful use of technology. For example, at least a 20% of representations should be ensured from civil society

organisations who have been raising issues of AI safety in India. The leadership of the AI Safety Institute should also have representatives from different governmental departments such as the Ministry of labour, IT, Health, Education etc. So that the specific concerns of these sectors are streamlined.

b) How can the Institute develop indigenous AI safety tools that are contextualised to India's unique challenges?

The Institute should support AI safety research and run regular schemes/platforms for third parties to test and report domestic and international AI systems and training models. Domestically established servers, cloud services, GPUs, and HPCs can also help assert data privacy and safety.

There has to be a mapping of unique safety concerns emerging from Indian contexts when AI technologies are deployed. For example, every innovation should have a separate safety committee that highlights the likelihood of issues that can arise from the deployment of an AI tool. It should have one external member (like an ICC committee at the workplace) with a proven track record in ethical AI practices. The AI safety institute should have a database of these external resources. The deployment of a technology should have a clearance from this committee.

c) Who should be the Institute's strategic partners?

A global-level collaboration through MoUs with various AI safety institutes such as the AISI (UK), AIST (USA), Center for AI Safety (San Francisco), CeRAI (IIT-Madras), and other public-backed and private AI Safety Institutes can be the first step for strategic partnership.

2. How can the AI Safety Institute build strong partnerships and gain stakeholder support?

a) What strategies will engage key stakeholders in supporting AI safety?

Offering a holistic platform for freelance and private AI systems developers and digital safety defenders and regularly organising engagements/competitions in AI safety could bring in talents to engage with the AI safety institute. Scholarships and research programmes on sub-themes such as AI ethics, human values, and AI, as well as training modules such as LMMs and LLMs, can also ensure private and public stakeholders become active participants in AI safety.

b) How can the Institute establish and maintain effective national and international partnerships?

Global AI Safety Framework engagements, collaborations between AI Safety Institutes and organisations, inter-governmental engagements in the form of MoUs, active participation in regulatory and policy discussions and events, etc., could build solid foundations for national and international partnerships for the institute. Dynamic collaborations with public and private players at the national and international levels are necessary to ensure the timely evolution of AI safety frameworks regarding technology and R&D.

International platforms like International Governance Forums can be used for having multi-stakeholder discussions. Organisations can apply for partnership based on the alignment with the mission of the institute.

c) What role should the Institute play in global AI safety discussions and standards?

The IndiaAI mission and AI Safety Institute shall actively try to become the global standard maker and forerunner in developing the collaborative framework in the domain. It should champion a model to mitigate the socio-economic risks and evolve the modalities. As a large user base for AI systems and products and a nation that

contributes highly to AI systems and training model developments, India could lead the discourses around Human Values and AI Ethics, AI User Awareness and Transparency and Explainability.

The institute should be able to streamline more local best practices as well as concerns in global forums.