COMMENTS

COMMENTS ON THE INITIAL DRAFT OF THE AI RISK MANAGEMENT FRAMEWORK

SUBMITTED BY – ANGSHUMAN KAUSHIK INDEPENDENT AI LAW AND POLICY RESEARCHER

COMMENTS

Dear esteemed members of NIST,

I shall take the liberty to put forward pithily, my comments for your perusal below.

I humbly request NIST to include guidance on the following;

- ***DEFINITIONS***

It is imperative to have clear definition(s) concerning the nomenclature used with respect to the domain of AI. Therefore, the definitions in the framework may be aligned with other legislations, code of ethics etc., pertaining to AI such as, the proposed AI Act of the European Union for ease and standardization.

- ***IMPACT ASSESMENTS***

As the AI systems severely impact individuals and certain sections of the society after deployment, hence, the framework may provide for the manner of conducting impact assessments such as, Human Rights Impact Assessments (HRIA) etc., before deployment. Further, the framework may also contemplate introducing something on the lines of 'protected class impact assessment' and the manner of conducting *ex ante* assessment of risks.

- ***POST-DEPLOYMENT MONITORING***

Constant monitoring of an AI system post-deployment is necessary to identify any risk not identified before or to keep new risks under check. It is also vital to track, review and reassess cases. The framework may provide for the manner of operationalizing post-deployment monitoring, review and reassessment of cases and *ex post* assessment of risks.

- ***RISK MANAGEMENT POLICY & ESTABLISHMENT OF COMMITTEES***

Before embarking on a journey to build an enterprise’s risk management policy with respect to AI systems, it is vital to know its 'risk appetite' and 'risk tolerance'. For instance, what should be done if a system crosses the enterprise's 'risk tolerance' level? Further, how, and at what stages, and mode and manner etc. risk should be evaluated. The framework may include recommendations on the above and also on oversight, auditing (both internal and external), documentation, internal controls for risk mitigation etc. It may also contain guidance on establishment of different committees to oversee whether the risk-related policies are implemented or not. The framework may also contain recommendations to the entities developing and deploying AI systems to refer to, risk incident databases, such as the one developed by the partnershiponai.org and study and implement, in appropriate and similar cases, the result(s) derived from those cases."

- ***PRIVACY AND CYBERSECURITY RISKS***

Although, privacy and cybersecurity risks are to be tackled by related frameworks, but there can be guidance on certain privacy and cybersecurity risks specific to AI systems, both at the design,

operational and the enterprise level and also with respect to aspects pertaining to allocation of risks, such as in the form of cybersecurity insurance.

- ***DIVERSE VOICES***

As general public is most likely to directly experience positive and adverse impacts of AI technologies, they should have the maximum involvement and say in its development. Hence, the framework may explore the possibility of offering guidance on the mode and manner and the stage(s) when diverse voices may be included in the development of an AI system. This will help in looking at risks not from the vantage point of a developer but from the perspective of the affected class and would give a better understanding of bias, fairness etc.

- ***TRAINING OF THE EMPLOYEES***

The framework may provide for the mode and manner of conducting regular training of the C-Suite and other staff on risk identification and mitigation techniques.

- ***NOTIFICATION OF RISKS(S)***

The framework may provide for the procedure to inform individual(s), groups etc., in case risk(s) concerning her/them are identified.

- ***MISCELLANEOUS ISSUES***

The framework may provide guidance to address risk posed by for example, non-representativeness in training datasets, 'drift', 'black box systems' in the form of 'explainibility', 'interpretability' etc. It may also offer recommendation on the incorporation of a mechanism whereby the moment an AI system enters into the 'risk-zone' i.e., a threshold is crossed, warning is triggered, and as a result, issues are flagged automatically. The framework may offer some guidance on the procedure for mitigating risks to the environment which has become imperative and is reflected in recommendations like the UNESCO Recommendation on AI Ethics.

- ***USE CASE – AN EXAMPLE***

A smart contract which is basically an algorithm, and run on a blockchain can have massive risks as these contracts do not have a human-in-the loop. There are other technologies also with which AI is either embedded or intertwined. The framework may provide for guidance on use cases like this or other appropriate cases.

Thanks

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx