RESPONSE TO NITI AAYOG'S DRAFT ON

RESPONSIBLE AI FOR ALL

SUBMISSION BY

The Dialogue

INFORM ENGAGE IDEATE

RESPONSE TO NITI AAYOG’S DRAFT ON

RESPONSIBLE AI FOR ALL

Submission by

The Dialogue

INFORM ENGAGE IDEATE

Authored by: Harsh Bajpai[1], Shruti Shreya[2], Trisha Pande[3]

Research Support: Antara Vats[4]

Editor: Kazim Rizvi[5]

Cover Illustration & Design: Abhinav Kashyap

[1] Doctoral Candidate and Part-TimeTutor at Durham University, U.K.
[2] Research Assistant at The Dialogue
[3] Policy Manager at The Dialogue
[4] Tech Policy Consultant at The Dialogue
[5] Founder at The Dialogue

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The **Working Document: Towards Responsible #AIforAll** by NITI Aayog has laid the groundwork for the development and deployment of ethical and responsible AI. The document has been successful in identifying core issues with the deployed use cases of Artificial Narrow Intelligence to screen candidates, detecting fraud in insurance claims and so on in tandem with constitutional values and the international standards. The principles of inclusivity and non-discrimination, equality, transparency, accountability, privacy and security, reliability and reinforcement of positive human values, laid down by the document are holistic and address the issues raised.

The Dialogue has prepared this report to assist NITI Aayog in arriving at a comprehensive framework for Responsible #AIforAll. The team has delved deep into each of the principles along with explaining the potential of AI and has identified the approach that can be adopted to uphold constitutional morality in the AI technologies. The direct and indirect impact of AI technologies has been further substantiated by taking into account the sociological dimensions as well.

List of Recommendations:
The team at The Dialogue has proposed the following recommendations Towards Responsible #AIforAll:

(a) *Increasing the focus on infrastructural development for research and innovation of AI use cases in the underdeveloped states to address the digital divide;*
(b) *Increasing representation in the data sets collected for modelling and training algorithms*
(c) *Increasing the frequency of independent audits for impact assessment;*
(d) *Ensuring transparency of the goals behind development of AI use cases through inclusion of policy makers and social scientists in the ethics committee;*
(e) *Ensuring environmental safeguards are in place to account for the environmental impact of developing AI technologies;*
(f) *Right to Explanation to be added under clause 17 in the Personal Data Protection Bill, 2019 to explain the use of personal data that is collected by the government or private entity ;*
(g) *Implement a Data Governance Framework for both Personal and Non-personal data.*

In this document, the team has also briefly outlined frameworks such as Algorithm Impact Assessment, Human Rights Impact Assessment and Trust and Fairness Framework which can aid NITI Aayog in building an inclusive framework that assists India in achieving #AIforAll.

INTRODUCTION

In the age of Artificial Intelligence governing our lives across sectors, NITI Aayog's Draft on *'Responsible AI for all'* is a timely development. There are a number of measures taken by the Government of India, primarily in dissemination of public welfare goods and services i.e. Aadhaar, which works on *'probabilistic algorithmic systems'*. As Frank Pasquale, in his book The Black Box Society states that these systems are inherently opaque in nature. In light of this, Niti Aayog's proposal of an Assessing Guide of AI systems is laudatory. Additionally, there are frameworks on the lines of transparency, accountability, fairness, human rights and sociological dimensions which have been recommended by *'The Dialogue'* team. These frameworks and specific recommendations would enable Niti Aayog to come up with a Working paper on 'Responsible AI'.

Evaluating the potential of AI in transforming economies, the Hon'ble Finance Minister in 2018-19 annual budget mandated NITI Aayog to establish a National Program on AI, with the intent to guide research and development in the emerging technologies in India. In pursuance of this goal, the NITI Aayog adopted a three pronged approach which included (a) undertaking exploratory proof of concepts of AI projects, (b) crafting a national strategy for building vibrant ecosystems, and (c) collaborating with various experts and stakeholders.

Since 2018 the NITI Aayog has been actively engaged in partnering itself with key technology players in the country for the implementation of AI projects in several public welfare sectors including health, agriculture and education. Learning from these projects it released a discussion paper in 2018 under the tagline #AIforALL which focussed on expanding the dimensions of India's social and economic growth.

Thereafter, in December 2019 a working document was released with the intent to develop an approach towards ensuring responsible usage of AI in India. This approach of "Responsible #AIforAll" was based on a consultation workshop facilitated by The Centre for the Fourth Industrial Revolution (C4IR), World Economic Forum. Following this, the working document was presented during a global consultation with AI experts around the world on 21 July 2020 and subsequently it was released by the NITI Aayog for wider public consultation.

It is pronged into five parts which analyse the principled approach envisaged by NITI Aayog. In our analysis we have utilised examples from India and foreign jurisdictions to better understand the practical implications of deploying the principles envisaged by the NITI Aayog. Where necessary, we have explicated the limitations in the application of the said principles and recommended measures/frameworks to ensure adherence. Lastly, the submission throws light on the '*privacy, ethics and gendered'* challenges in the consultation paper while suggesting implementable '*Technical and Human Rights Impact Assessment'* frameworks to counter these issues effectively.

1. CONSTITUTIONAL DUTIES AND CONSTITUTIONAL MORALITY OF A RESPONSIBLE AI

1.1 Principle of Inclusivity and Non-discrimination

Slide 29 of NITI Aayog draft enumerates principle of inclusivity and non discrimination to be integral for furthering the non harm principles and building a Responsible AI.

In the case of *State of West Bengal v. Anwar Ali Sarkar*[6] the Hon'ble Supreme Court has recognised Non Discrimination to be the Heart of Republicanism. However, despite the judiciary recognising the significance of this principle, technology continues to remain a privilege in the hands of the socio- economic affluent classes. The online classes in times of COVID19 and the innumerable stories of people being compelled to sell off their assets to buy laptops for their children or students sitting on rooftops to get uninterrupted broadband signal is an embezzling evidence of this fact.[7] According to the report by Telecom Regulatory Authority of India in 2018, internet penetration in India is about 49% and within that only 25% people belong to rural areas[8]. Lack of infrastructural resources and erratic power cuts further complicate the issues of technological access in states such as Assam, Himachal Pradesh and Jammu & Kashmir. Since most AI systems collect data from publicly available datasets, the digital divide would further aggravate the issue of exclusion and discrimnation due to the non-representative data sets that will be generated for coding and testing of the algorithms. This might further add onto the plight of the vulnerable and underrepresented populations including the less educated, low skilled, women and differently abled persons[9].

Equitable distribution of resources and opportunities among all the sections of the society and across the entire geographical boundary of India is an essential component of Right to Equality under Article 14 and Right Against Non Discrimination under Article 15 and 16 of the Indian Constitution. The said principle is also a part of Directive Principles of State Policy under Part

[6] AIR 1952 SC 75.

[7] Online Classes: Poor Students in Delhi Struggle due to Lack of Internet Connections, New Indian Express, https://www.newindianexpress.com/cities/delhi/2020/may/18/online-classes-poor-students-in-delhi-struggle-due-to-lack-of-internet-connections-2144781.html.

[8] Telecom Regulatory Authority of India (2018), *The Indian Telecom Services Performance Indicators*. https://www.trai.gov.in/sites/default/files/PIR08012019.pdf.

[9] UK (2018), *AI Sector Deal*, Department for Business, Energy & Industrial Strategy and Department for Digital, Culture, Media & Sport, Government of the United Kingdom, https://www.gov.uk/government/publications/artificial-intelligence-sector-deal.

IV of the constitution where the state ought to strive to achieve social, economic and political justice for all its citizens under Article 38 and 39 of the Constitution.

In order to successfully attain this goal it is imperative that our AI policies are inclusive of the interests of all genders, religious and social communities. Furthermore, it is imperative to design AI applications that aim at the upliftment of the marginalised sections. For instance, designing applications like the Seeing AI mobile application of Microsoft which scans and recognises everything around and gives an oral description for the benefit of the visually impaired people.[10]

Likewise, the principle of Cooperative Federalism must be applied wherein state and local self government bodies should work in collaboration with the central government to ensure even application of the AI policies and applications till the grass root level. In order to ensure adequate representation of the backward regions, preference should be given to establishing AI Centres of Excellence in the less developed states of the country. Such initiatives would help in better penetration of technology to the interiors of the country and also aid the creation of job opportunities in these states.

Recommendations:
(a) *Designing specific AI applications for the aid and welfare of underprivileged communities;*
(b) *Ensuring equitable dissemination of technological resources and training and awareness campaigns about latest technological developments in an easily inferable manner for the marginalized communities;*
(c) *furtherance of cooperative federalism goals and establishing AI Centers of Excellence in the underdeveloped regions of the country.*

1.2 Principle of Equality

Another important facet of making a 'Harmless Responsible AI' as enumerated in **Slide 28** of the present draft is to ensure that the benefit of AI technology is 'equally' disseminated across all the sections of the society. Moreover, it implies the application of the principle of non

[10] Villani, C. (2018), *For a Meaningful Artificial Intelligence - Towards a French and European Strategy*, AI for Humanity, https://www.aiforhumanity.fr/.

arbitrariness which is an essential facet of Right to Equality as enunciated by the Apex Court in *E.P. Royappa v. State of Tamil Nadu.*[11] This principle is enrooted in the Aristotalian theory of distributive justice which entails finding means between two extremes and ensuring justice for each facet of the system.[12] However, in the recent years technological expansion at the behest of human and ecological value has been rampant. Autonomous sonars used for drilling oil in Bombay High have been one of the primary reasons for decrease in marine biodiversity by almost 25% in the past ten years.[13]

Right to Equality has been held to be a part of Basic Structure of the Indian Constitution in the landmark judgement of *Kesavananda Bharati v. Union of India.*[14] However, the grund norm envisaging this principle is not a blanket approach to equality wherein every person is meted out the same treatment. Rather it implies equality based on the principle of reasonable classification wherein the deprived classes are entitled to affirmative actions to ensure their socio-economic upliftment. Thus the entities which are supposed to be treated the same and the ones that aren't, need to be determined on a case to case basis and ensuring that such a determination is done accurately by an AI system pausits the real challenge.[15]

Recommendations:

(a) *Following the sustainable development principles during research and deployment of technology;*

(b) *Creation of bias free algorithms;*

(c) *Subjecting the training data sets to multiple rounds of testing and auditing using authoritative tools, and self-regulatory or regulatory approaches. For example, algorithmic impact assessments in predictive policing systems.*[16]

[11] AIR 1974 SC 555.

[12] Copp, D. (1992). The Right to an Adequate Standard of Living: Justice, Autonomy, and the Basic Needs. *Social Philosophy and Policy*, *9*(1), 231-261.

[13] Mardani, Gashtasb, et al. "Application of Genetically Engineered Dioxygenase Producing Pseudomonas putida on Decomposition of Oil from Spiked Soil." *Jundishapur Journal of Natural Pharmaceutical Products* 12.3 (Supp) (2017).

[14] AIR 1973 SC 1461.

[15] Muller, C. (2017), *Opinion on the Societal Impact of AI*, European Economic and Social Committee, Brussels, https://www.eesc.europa.eu/en/our-work/opinions-information- reports/opinions/artificial-intelligence.

[16] Price, A. (2018), "First international standards committee for entire AI ecosystem", *IE e-tech,* Issue 03, https://iecetech.org/Technical-Committees/2018-03/First-International-Standards- committee-for-entire-AI-ecosystem.

1.3 Principle of Transparency and Accountability

The third principle under **slide 28** is the Principle of Transparency and accountability. In the words of Justice Madan B. Lokur *"The balance between transparency and accountability is crucial for building a privacy protecting state".*[17]

Both the present draft of NITI Aayog and even the 2018 report has explicit mention of this principle. Still, a lot of scope remains before the full realisation of these values in the Indian technology discourse. For instance, in the Aarogya Setu App the source code has been made public which is indeed a welcome step towards creating a privacy respecting setup. However, several other measures may be taken to further ensure transparency and accountability in the functioning of the app.[18] Some of these measures include, releasing a manifesto or a website that would mention the details of the project and its purposes thereof and the revelation of the server code of the app to help people gather insights into the exact protection and regulation scheme of the main data center.

The Principle of Transparency should not just encompass how the AI is used in prediction, recommendation and decision but also how it is developed, trained and deployed along with the internal factors that determine its reaction.

Similarly, the Principle of Accountability should not just take into account the AI systems but also the decisions made by the AI systems and also the organisation that is using or funding the development of that AI. Inspiration can be taken in this regard from the New York City's proposed Algorithmic Accountability Act that seeks to frame accountability of individuals, organisations and processes rather than the AI system itself.

Experts at Harvard University in the Berkman Klein Center Working Group on Explanation and the Law have identified approaches to improve transparency and accountability of AI systems, which is as follows:

[17] ABC v. State of Delhi, 2015 SCC OnLine SC 609.
[18] Privacy Framework for the Aarogya Setu App, The Dialogue, Working Paper, Version 1.0, https://thedialogue.co/wp-content/uploads/2020/05/Privacy-Framework-for-the-Aarogya-Set-App.pdf?fbclid=IwAR0u0FqyxdXfzmKtV-BhTweVqNLVyng8bRfJXVe1DgACKd8vqW4NlKo2FUQ.

Approach	Description	Well suited Context	Poorly Suited Context
Theoretical Guarantees	In some situations, it is possible to give theoretical guarantees about an AI system backed by proof.	The environment is fully observable and both the problem and solution can be formalised.	The situation cannot be clearly specified (most real world settings).
Statistical Evidence Probability	Empirical evidence measures a system's overall performance, demonstrating the value or harm of the system without explaining the specific decisions.	Outcomes can be fully formalised; it is acceptable to wait to see negative outcomes to measure them; issues may only be visible in aggregate.	The objectives cannot be fully formalised; blame or innocence can be assigned for a particular decision.
Explanation	Humans can interpret information about the logic by which a system took a particular set of inputs and reached a particular conclusion.	Problems are incompletely specified, objectives are not clear and inputs could be erroneous	Other forms of accountability are possible.

Figure 1: Doshi-Velez et. al. (2017) "Accountability of AI under the law: The role of explanation", https://arxiv.org/pdf/1711.01134.pdf

Recommendations:

(a) *Due disclosure of server code of apps and utilities deployed in public functions;*

(b) *ensuring verifiability of input and output data;*

(c) *transparency about the goals of an AI system and about its results;*

(d) *awareness and understanding about the reasoning process of AI;*

(e) *equitable accountability of the organisations and individuals using the AI system.*

1.4 Principle of Privacy and Security

In the words of Justice Madan Mohan Sapre, *"The right to privacy of an individual is essentially a natural right, it is a right with which a person is born and remains with him/her till he/she breathes his/ her last. In short, privacy is inalienable and inseparable from human beings."*[19]

Right to Privacy has been recognised as a fundamental right under Art. 21 of the Indian Constitution in the case of *Justice K.S. Puttaswamy (Retd.) v. Union of India.*[20] Further, a three fold test has been laid down whereby any infringement into this right of the people must satisfy the principles of necessity, legality and proportionality. The significance of these principles in order to safeguard a constitutionally protected right has also been elaborated in the cases of *Modern Dental College and Research Centre v. State of Madhya Pradesh*[21] *and Om Kumar v. Union of India.*[22]

In pursuance of this constitutional mandate, the present NITI Aayog draft recognises Privacy and Safety to be essential facets of a Responsible AI in **Slide 28**. Despite the elapse of 3 years since the Puttaswamy judgement we are still in the trial stage and need to put in more efforts to have an effective data protection regime. However, the collection of personal data at every stage remains rampant be it on social networking sites or while booking travel tickets to something as small as placing a food order on a food ordering app.[23]

It is suggested that collecting more data about an entity or enterprise allows one to make a more accurate prediction. However, this exercise leads to a data privacy paradox wherein, while the higher amount of data collected reduces the risk of bias from a skewed sample but the risk of privacy and security gets amplified.[24] Therefore, in order to counter this challenge it is paramount to put in place adequate privacy ensuring safeguards especially when one is dealing with personal data. The 2019 OECD privacy guidelines for data privacy can be pertinent in this regard. Though the present PDP Bill, 2019 takes into account principles like collection limitation (including, where appropriate, consent as means to ensure this principle), purpose

[19] Justice K.S. Puttaswamy v. Union of India, 2017 10 S.C.C. 1.
[20] Id.
[21] AIR 2016 SC 2601.
[22] 2001 (2) SCC 386.
[23] Ferracane, M. F., Kren, J., & van der Marel, E. (2020). Do data policy restrictions impact the productivity performance of firms and industries?. *Review of International Economics*, *28*(3), 676-722.
[24] Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security*, *64*, 122-134.

specification, but need more safeguards in the form of openness, individual participation in the form of bug bounty programmes and accountability.[25]

In addition to these principles, ensuring anonymisation of personal data, use of cryptographic tools for data processing, decentralised storage and most importantly having separate legislations for governing both personal and non personal data along with an independent data protection authority is cardinal.

Recommendations:
(a) *Enactment of the Personal Data Protection Bill, 2019 at the earliest;*
(b) *designing a robust privacy framework for Non Personal Data;*
(c) *ensuring that these legislations conform to the principles of legality, necessity and proportionality;*
(d) *having a strong independent and statutorily recognised data protection authority that is free from state interventions.*

1.5 Principle of Safety and Reliability

As research in the field of technology expands there exists a growing concern about safety of the research subjects and users of the invented technology. This is the reason the draft in **slide 28** recognises safety and reliability of the research as a separate principle which is critical in the policy framework of a Responsible AI.

To ensure that the machines developed are safe and reliable it is important that they are robust and stable in their performance and no harm is caused to the life or body of the user or any third party during its usage. Presently, the safety impact of many AI technologies seems uncertain especially the AI deployed in the public health sector. Companies like IBM have been designing AI technologies to leverage research and diagnosis in the health sector. However, many reports have surfaced highlighting the inaccuracy in the AI diagnosis, thus putting the reliability of the technology into question.[26]

[25] OECD (2019), *Recommendation of the Council on Artificial Intelligence*, OECD, Paris.
[26] Yang, Yi, et al. "The diagnostic accuracy of artificial intelligence in thoracic diseases: A protocol for systematic review and meta-analysis." *Medicine* 99.7 (2020).

Researchers in China have developed a Fault Tree Analysis Model which is gaining prominence in terms of easily understandable graphical features.[27] Under this model machine learning and real time operational data are used to know the normal behaviour of the machines. Following this, if any abnormal behaviour is observed in the machines then using this approach, abnormality is determined on the fault tree and then the operator is informed accordingly.[28]

Recommendations:
(a) *Adequate testing of the safety and accuracy of an AI before deployment;*
(b) *creation of robust and stable systems that are able to withstand boisterous conditions;*
(c) *deploying easily comprehensible safety models (like the Fault Tree Analysis Model that uses graphics) using which even a novice can easily understand the faults in a machine.*

1.6 Principle of Reinforcement of Positive Human Values

India's strong ethos of fundamental rights mentioned in the Constitution if incorporated and implemented well in its AI policies, holds the potential to strengthen its reputation in the international community and set us apart from the countries that seek to further authoritarian narratives around technological governance.

To this end, it is important that there is fair dissemination of ethical and human rights values in our research and technology. The NITI Aayog draft in **Slide 27** acknowledges the ethical values and standards prescribed by international organisations like United Nations Educational, Scientific and Cultural Organisation (UNESCO) and Institute of Electrical and Electronics Engineers (IEEE) and recognises the need to have standard setting bodies that would ensure that our technology policies are compatible with these standards. However, proper guidelines and benchmarks need to be created for the functioning of this body along with the composition and requisite qualification and skills of its members.

The type of ethical code furthering human values that must be programmed in a particular machine needs to be decided on the basis of the user and the usage of the machine. For instance, an AI that is deployed during a pandemic to decide upon resource allocation needs to be feeded the algorithmic codes that further the principle of utilitarianism (maximum benefit for maximum people).[29] On the other hand, supposedly a terrorist attack takes place

[27] Koorosh Aslansefat, IEEE Xplore, Safety+ AI: A Novel Method to Update Safety Methods Using Artificial Intelligence (2019),
[28] Id.
[29] Müller, V. C. (2020). Ethics of artificial intelligence and robotics.

and the CCTV camera only captures the image of the criminal's eyes as he was wearing a mask. So now to trace the terrorist personal data of all the known terrorists is feeded into an AI. The 'means' used, that is to feed in personal data of so many people into an AI is not privacy respecting but the end goal of this act is justified.[30] Therefore, the AI deployed for this task should be programmed with the algorithmic codes of consequentialism (dynamic logical codes)[31] that helps it to reason about the consequences of its actions and operate .

Recommendations:

(a) *Establishing independent autonomous bodies to determine standards for the research and enforcement of new technology policies in India;*

(b) *Ensuring that these standard setting bodies are presided over by competent data scientists and human rights experts;*

(c) *undertaking periodic human rights impact assessment.*

[30] Hakli, R., & Mäkelä, P. (2019). Moral responsibility of robots and hybrid agents. *The Monist*, *102*(2), 259-275.

[31] Mexico (2018), *Towards an AI strategy in Mexico: Harnessing the AI Revolution*, British Embassy Mexico City, Oxford Insights, C minds, http://go.wizeline.com/rs/571-SRN- 279/images/Towards-an-AI-strategy-in-Mexico.pdf.

2. ANALYSING THE POTENTIAL OF AN AI

2.1 The Concept of 'AI Garage'

The present draft in **Slide 3** envisions India as an AI garage. However, before moving ahead with the policy, robust cybersecurity infrastructure and effective legislative framework for AI and data protection needs to be put in place.

The model of AI garage was first conceived by NITI Aayog in its National Strategy Document. It envisioned India to export Artificial Intelligence as a service (AIaaS). The report culled out a wide range of sectors where Indian can provide large incremental value i.e. Healthcare, Agriculture, Smart Mobility, Retail, Manufacturing, Energy, Smart Cities, Education and Skilling. However, it should not mean that poor, rural, marginalized people are made the guinea pigs for testing of the AI systems manufactured and exported to the world. As outlined in sections below, India lacks a privacy, ethical and regulatory framework without which the concept of AI garage is scratchy at its best.

It should be noted herein, that there are several countries aiming to make themselves as one stop solutions service providers for AI systems. However, to become one, not only the institutional structure but the means to realise the ambitions should be panned out too. Further, the policy goal should be in such a nature that by making India a data marketplace we do not end up creating domestic oligopolies at the cost of smaller Indian players. Thus, India needs a coherent framework across sectors to emerge as a powerhouse:

1. In the H-Index ranking (a measure of how often its papers are cited)[32] India ranks 9th in the world. Of this, if papers related to emerging technology are considered, India ranks 18th in the world. Thus, primarily it is the quality of indigienuous research across sectors which is of paramount importance. Rather than adopting AI technologies in healthcare and education which have inherent biases, especially towards countries with brown skin, India's 'Make in India' research would prove useful.

2. It is imperative to pay those workers who are training the AI systems. The jobs of annotating AI datasets is often underpaid coupled up with poor employment

[32] International Science Ranking, https://www.scimagojr.com/countryrank.php, accessed 25th Aug, 2020.

conditions. Exploring how artificial is an Artificial Intelligence system shows that there are thousands of online curators, trainers and responders behind the AI curtain.[33] In the future humans in the AI loop system would work where present day occupations would be transformed with tech innovation services. This workforce needs training, support and compensation for a proper working environment to work and enhance AI systems.

3. AI systems are increasingly getting controlled by few. Even if smaller players are using AI systems, most of the datasets are acquired from the few global players. This creates *'data network effects'* quickly leading to certain companies building up proprietary dataset dominance in the market. It results in specific AI solutions and recommendations which overtime turn into best industry practices. This not only leads to biasness, non-inclusivity and discrimination but also builds a bad set of 'best practices'. Thus, before investments, there should be a Human rights impact assessment as outlined in our 4th chapter herein.

Recommendations:

(a) *A privacy, ethical and regulatory framework is much needed;*

(b) *Subsidies to AI startups through the Technology Development Programme without any biases;*

(c) *Improvement in the research quality by promoting peer review publications and innovation as it will boost soft power across sectors;*

(d) *Proper aid package to the contractual AI trainers in terms of amending labour laws and social security schemes;*

(e) *All investments should be subjected to the Human Rights Impact Assessment approach;*

(f) *Declaring all AI storehouses as critical information infrastructure under Section 70 A of the Information Technology Act thereby putting them under the direct regulation of the Computer Emergency Research Team (CERT). However, this declaration should not bar Right to Information requests.*

[33] Mary L. Grat & Siddharth Suri, "The Humans working behind the AI 19 curtain", Harvard Business Review, 9 January, 2017.

2.2 Usability of AI Technologies

There has been an upsurge of use cases of AI technologies all over the world in numerous sectors like education, transportation, agriculture, advertising, policing, finance, marketing, healthcare and so on. In India, the government is investing in the adoption and deployment of AI technologies in maintaining grain value chains and land records, managing traffic on the roads, forecasting weather updates, predicting health diseases, and for enhancing educational outcomes amongst other use cases.

The government is also envisioning AI as a tool for prediction farming which can aid in doubling the income of the farmer by ensuring the security of the crop sown. In the report by NASSCOM (2020), 'Unlocking Value from Data and AI: The India Opportunity'[34], the organisation has claimed about 45% of visionary value i.e. the value generated from AI technologies will be through the consumer goods retail, agriculture, banking and insurance sectors.

The government has also been harnessing the power of technology to enhance quality of life and driving economic growth through the Smart Cities Mission. One such example is The Pune Street Light Project which is set up to ensure energy efficient usage of street lights which would be controlled remotely through a Supervisory Control and Data Acquisition (SCADA) system. India is planning to augment the use of AI in smart cities to build smart parks, smart rooftops and for transportation within the city. The widespread use of AI technologies is influenced by outcomes that improve efficiency, productivity, decision making and risk management strategies. But, along with the aforementioned factors, determinants like lack of technical infrastructure in cities, structured data, regulatory barriers, privacy considerations, and ethical issues also play a huge role in determining the extent of usability.

Due to the existing digital divide between states, genders, and social classes, the data generated evidently leads to certain outcomes which are discriminatory in nature especially in the cases of hiring.

In 2019, the Ministry of Electronics and Information Technology developed a draft report titled 'On Platforms and Data on Artificial Intelligence'. The report recommends the creation of a National Artificial Intelligence (AI) Resource Platform (NAIRP) for India, to create a common

[34] NASSCOM (2020), 'Unlocking Value from Data and AI: The India Opportunity' https://www.nasscom.in/knowledge-center/publications/unlocking-value-data-and-ai-india-opportunity

platform for those interested in using AI for societal good.[35] It mentions that "the platform will also have scope for sharing and driving standards, policy guidelines, entrepreneurship and developing a creative economy." This platform is modelled after IIT Kharagpur's 'National Digital Library of India (NDLI) Project'[36] and will serve as a public open platform. The timeline for the project from its start is expected to be 3-4 years, with an estimate of about Rs. 100 crores. The report mentions that there is a need to conduct an extensive gap analysis on open data policies, and a network of partners will be sought for that purpose.

Recommendations:

(a) *The present draft has outlined certain use cases and the issues it raises, but in order to develop anticipatory policies to take into account all the principles, the government has to conduct an Algorithmic Assessment before deploying each of the use cases. In order to ensure a holistic impact assessment, the ethics committee should also include policymakers;*

(b) *While the guidelines for Responsible AI take into consideration the global standards and global frameworks, it has to take into account the context of the society and the various intersecting social realities that lead to generating skewed datasets;*

(c) *The funding for Research and Development of AI use cases should mandatorily be accompanied with research on AI ethics and safety;*

(d) *Whilst conducting the gap analysis for NAIRP, it would be useful to include members from civil society in the consultations - especially intersectional researchers working across public policy, humanities and engineering.*

2.3 Impact of AI on 'Work'

Both the present draft and the NITI Aayog report of June 2018 consider AI as a societal consideration as it can disrupt jobs. In 2003, Autor, Levy and Murnane (ALM Model)[37] focused on routine and non-routine tasks and the degree to which physical and cognitive work required respectively. Their study concluded that the routine tasks (which consist of both manual and cognitive tasks) can be substituted by computer capital as they are programmable in nature. The non routine tasks which include skills like perception, problem-solving and intuition can at maximum be complemented by AI systems. However, the ALM

[35] A Report on Platforms, MEITY, https://www.meity.gov.in/writereaddata/files/Committes_A-Report_on_Platforms.pdf, accessed Aug 25th 2020.
[36] National Digital Library of India, MHRD, https://ndl.iitkgp.ac.in accessed Aug 25th 2020.
[37] David H. Autor, Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," Quarterly Journal of Economics 118, no. 4 (2003): 1279–333.

model has become complicated now due to AI systems being able to comprehend past use cases and blurring the line between routine and non-routine cases. For e.g. in the ALM model, truck driving was categorised as a nonroutine manual work but now driverless trucks are in the market[38] and the industry is moving towards being fully autonomous. The key limitation to ALM model or Frank Levy's or Frey and Osborne methodology is that none of them while forecasting automatability took account of variables like, internal organization, institutional and regulatory landscape or degree of unionization. Issues regarding race, gender, votability and market power of marginalised communities etc, demographics, geographic location, lifestyle patterns etc.. would come into the picture while automating a particular sector, job or task.[39] For e.g. due to increase in retail online shopping, brick and mortar stores are on a decline and the jobs like departmental store associates would not be needed. However, it cannot be regarded as automation of a job as a similar skilled job is available on the *'intelligent supply chain'* of an online store.

The present draft raises the issue of societal consideration of technology displacing workers by automating jobs. The report also highlights solutions such as skilling, adapting legislations and regulations to leverage benefits and harness new job opportunities. However these are broad terms and do not cover the entire gamut of issues surrounding the ecosystem. Therefore we propose a framework encompassing the policy issues to be fixed across sectors in order to be ready for the coming generation jobs. The framework includes both who are in the workforce or who are yet to join it:

1. **Business - Govt. Partnership:** Different sectors which are facing the brunt of automation should engage with the government and the academic institutions to provide advanced educational and up-skilling opportunities to the upcoming and the existing workforce. There is a need to develop a cultural shift in the Indian education system by instilling innovation and cross-sectional education at the center. The education scenario in India is further worsened due to the huge literacy and education gap which obviates women from learning social, interpersonal, soft and entrepreneurial skills. Adding to this is the backlash from the patriarchal setup, especially the high caste men who see technological enablement of women as a threat to established familial and societal structure.

[38] Steve Viscelli, "Driverless? Autonomous Trucks and the Future of the American Trucker," Center for Labor Research and Education, University of California, Berkeley, and Working Partnerships USA, September 2018.
[39] Moradi, Pegah. "Race, Ethnicity, and the Future of Work." (2019).

IIT Delhi's 'PhD Incubator Program' to allow PhD students to kickstart a startup instead of a thesis is a move in the positive direction. This program not only boosts technological disruptions in academic institutions, but also generates employment opportunities.[40]

2. **Education Policy:** The National Education Policy (NEP) 2020 conceptualises the National Educational Technology Forum (NETF), as a "a platform for the free exchange of ideas on the use of technology to enhance learning, assessment, planning, administration, and so on, both for school and higher education".[41] One of the reasons for the NETF's existence, as outlined in the policy, is to identify the decline in jobs because of disruptive technologies such as AI. The NETF will alert the Ministry of Education by way of periodic analysis which categorizes the technology and its timeframe for disruption. However, without public investment in underlying services such as telecom infrastructure in unconnected areas, the policy will by design exacerbate the digital divide.[42] Whilst the policy addresses the importance of skilling students in upcoming areas such as AI, by including this as a contemporary subject, the underlying issue is that many students will not have access to the requisite Information and Communication Technology (ICT) required to learn the curriculum. The outdated form of learning about computer systems through printed textbooks is still followed by state boards and the Central Board of Secondary Education (CBSE). If the policy expects to bridge this gap, then it should also explicitly come out to address the digital divide - in terms of caste, gender, class, ability - and so forth. Any analysis will otherwise reiterate the same findings - that those at the margins of society are left out from the benefits ICT has to offer for learning AI and upskilling students.

3. **Curriculum Framework:** Given that the NEP 2020 has a focus on equipping school students with coding abilities and emphasizing the importance of learning AI as a life-skill, especially in higher education institutions, it is necessary to include an AI+Ethics module in the curriculum of schools and colleges. An ideal curriculum will break out of the pattern of rote learning code, and teach students how to think about the different

[40]IIT Delhi, PhD incubator, Call for Applications, Accessed August 18th, 2019, http://www.iitd.ac.in/content/iitd%E2%80%99s-phd-incubator-programme-call-applications-0.
[41] https://www.mhrd.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
[42] A Dictionary of National Education Policy 2020 a handbook from COLLECTIVE August 2020.

stakeholders in the AI ecosystem and what needs each stakeholder has. Students must be able to understand the impact of AI on societal realities.[43]

4. **Skilling Ecosystem:** Those who are already in the workforce are highly dependent on their employer for training and upskilling purposes. It is incumbent upon the professional to kindle the curiosity and creativity within. While the companies that were looked into during the research incorporate skilling programs, a need for incentivising demand-driven skilling programs within the larger ambit of government skilling programs has been felt.[44] While this was the idea behind underpinning the establishment of Sector Skill Councils, the Report of the Committee for Rationalization & Optimization of the Functioning of the Sector Skill Councils highlights several gaps plaguing the SSC framework.[45]

   Platforms that allow the industry to work in tandem with the government could be one approach. This would entail utilising the industry's capital prowess and technological know-how to aid in the development of skills that are anticipated to be required in the future of work within the IT sector, and subsuming the industry's strengths into government envisioned skilling programs.[46] Another financing model that is receiving consideration is a 'Reimbursable Industry Contribution' which would be a common pool of industry contributions to be utilised towards the training of manpower.[47]

5. **Social Security:** Building from the assumption that social security is a fundamental part of the implicit social contract of modern societies, it then becomes crucial to widen the net of social protection mechanisms by incorporating the realities of the future of work. What would be required is a fundamental shift in how social protection systems are currently premised on an existing unique employer-employee relationship. It also then becomes crucial to envision ways in which the employer-employee relationships are seen. The ILO's Recommendation No. 198 developed in 2006 at the General

[43] AI + Ethics Curriculum for Middle School, https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/, accessed Aug 24th, 2020.
[44] Report of Future of Jobs, NASSCOM, http://ficci.in/spdocument/22951/FICCI-NASSCOM-EY-Report_Future-of-Jobs.pdf,accessed Aug 24th, 2020.
[45] Shri Sharda Prasad et al., "Report of the Committee for Rationalization & Optimization of the Functioning of the Sector Skill Councils," Ministry of Skill Development, Last modified December 2016, https://www.msde.gov.in/assets/images/ssc-reports/SSC%20Vol%20I.pdf
[46] Anurag Malik, Future of Jobs in India- A 2022 Perspective, NASSCOM & Ernst and Young, Accessed August 4th , 2019, http://ficci.in/spdocument/22951/FICCI-NASSCOM-EY-Report_Future-of-Jobs.pdf.
[47] Report of the Committee for Rationalization and Optimization of the Functioning of the Sector Skill Council, December 2016, Volume I, Accessed August 2nd 2019, https://www.msde.gov.in/assets/images/ssc-reports/SSC%20Vol%20I.pdf.

Conference can be a very useful starting point. It emphasises on understanding working arrangements basis the nature of the work that is being carried out as opposed to how the arrangement is described contractually.

6. **Legal Needs:** There should be an emergence of scheduling laws which require managers to announce schedules further in advance, end shifts and create minimum shift lengths. Similarly, the Ministry of Labour should come up with a law that classifies gig workers as employees too so that they are not shred of all labour rights (like minimum wage, unionization etc.). A different law towards gig workers can also be framed which articulates compensation regimes accurately recognizing workers time and effort.

Recommendations: *Skilling is not the only requirement to train our workforce and prepare for the future jobs. Some other factors which the government can take into consideration are:*
(a) *Enabling lifelong learning capabilities;*
(b) *Incentivising industry involvement in skilling;*
(c) *Strengthening Social Protection;*
(d) *Building an R&D ecosystem;*
(e) *Prepare scheduling laws for the blue-collar workers;*
(f) *Special protection for 'gig economy'.*
(g) *National Curriculum Framework (NCF 2020) is in the making and should be developed in such a manner that it enables lifelong learning capabilities.*
(h) *Include an AI+ Ethics module in schools and colleges*

2.4 AI and Network Infrastructure in India

NITI Aayog's **National Strategy for Artificial Intelligence** discussion paper emphasizes the initiatives of other countries to provide supply-side interventions of infrastructure, in order to develop a larger Artificial Intelligence (AI) ecosystem. Governments have focused on digital connectivity infrastructure, comprising 5G/full fibre networks and fiscal incentives, amongst other initiatives.[48] Similarly, the **5G Steering Committee's report** discusses the need to develop a base of core technologies such as computing, communications, artificial intelligence, signal processing, security, and block chain. These core technologies will build a robust ICT system which will pave the way for new technologies such as 5G. The number and

[48] NITI Aayog (2018). *National Strategy for Artificial Intelligence*. New Delhi.https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

variety of 5G links expected to permeate India will increase by 100 times more than the existing links supported by 2G/3G/4G networks. Thus, the Steering Committee anticipates that machine learning and artificial intelligence will play a central role as 5G networks develop. This will be of benefit to operators, since they will be able to evolve their business models, provide services at lower price points; and overall offer enhanced performance and reliability in terms of networks.[49]

1. India's Infrastructural Requirements: The mission 'Propel India' envisaged in the NDCP 2018 aims to promote **investment, innovation and Intellectual Property Rights (IPR)** - which will be helpful in promoting Industrial Revolution 4.0 by harnessing the power of technologies such as 5G, Artificial Intelligence (AI), Internet of things (IoT), Cloud and Big Data.[50] Currently, the telecom industry has about 471,000 towers in India, and is expected to see at least 100,000 more towers with an investment of $2.78 billion towards the Digital India initiative.[51] Telecom operators are presently investing in small cells, fibre networks, in-building solutions, street furniture, and Wi-Fi hotspots which will densify the existing networks and make the country ready for the deployment of 5G. Infrastructure providers are expected to follow suit, and initiatives such as Digital India, BharatNet and Smart Cities have addressed the necessity of telecom towers for their smooth functioning.

2. Infrastructure and Privacy - A Prerequisite for AI: Privacy by design is becoming the norm in most products and services, and it is finding its way into data protection regimes created by governments across the globe. It is centred around the philosophy that risk which can arise from digitalization should be anticipated by companies and user privacy should be protected throughout the product/solution life cycle. Governments which are creating laws around data privacy and personal data are including these principles in the law. For instance, in India's **Personal Data Protection Bill 2019, privacy by design** has been addressed. Since telecom companies are providers of technical solutions and play a major role in the digital value chain, there is added emphasis on them to preserve privacy across the services they provide. The reason telecom operators play this significant role is because of their unique ability to act as

[49] Dot.gov.in. (2020). *Report of the 5G High Level Forum*. [online] Available at: https://dot.gov.in/sites/default/files/5G%20Steering%20Committee%20report%20v%2026.pdf?download=1 [Accessed 1 Mar. 2020].
[50] https://main.trai.gov.in/. (2019). *Pre-Consultation Paper on Enabling Unbundling of Different Layers Through Differential Licensing*. [online] Available at: https://main.trai.gov.in/sites/default/files/CP_09122019.pdf [Accessed 1 Mar. 2020].
[51] https://www2.deloitte.com/in/en.html. (n.d.). *5G: The Catalyst to Digital Revolution in India*. [online] Available at: https://www2.deloitte.com/content/dam/Deloitte/in/Documents/technology-media-telecommunications/in-tmt-the-catalyst-report-one-noexp.pdf [Accessed 1 Mar. 2020].

the intermediary between businesses and consumers. Therefore, they can also potentially leverage this position to improve their trustworthiness and garner investment to work towards building secure networks.

Recommendations:
*In the way that the network infrastructure laid out by telecom operators will benefit AI, the reverse will also be a reality in the near future. As the scale, **complexity and continued growth** forecast automation in the telecommunications industry (since telecom networks are expected to grow by about 10,000 times of their existing scope) - which will require intelligent automation technologies to improve operating costs.*

***(a)** It will allow operators to provide an enhanced **speed of service and quality of service** for customers.[52] With heavy consumption of Over-the-Top (OTT) content in the digital age (for example, video streaming platforms) there is a high demand for bandwidth. This creates financial and operational pressure on telecom operators, which are functioning in a stressed financial state at present;*

***(b)** The solution AI offers to the telecom industry is to reduce its **Operating Expenditure (OpEx) on manual configuration and service provision tasks**; given that the industry's **Capital Expenditure (CapEx) is already significant (for example, investment in data centres, towers etc.)**.[53] According to a report by the credit rating agency ICRA, increasing data requirements have increased the average CapEx for telecom operators between 2017-19 to more than ₹1 lakh crore in FY19.[54] Therefore, the deployment of AI can help significantly in uplifting the financially stressed condition of the telecom sector.*

[52] www.infosys.com. (2017). *Intelligence is in the Airwaves for Telecoms Firms*. [online] Available at: https://www.infosys.com/human-amplification/documents/telecommunications-ai-perspective.pdf [Accessed 1 Mar. 2020].

[53] *Ibid at 6*

[54] EconomicTimes.IndiaTimes. (2019). *Telcos' Peak 4G Capex Cycle is Over: ICRA*. [online] Available at: https://economictimes.indiatimes.com/industry/telecom/telecom-news/telcos-peak-4g-capex-cycle-is-over-icra/articleshow/71367449.cms?from=mdr [Accessed 1 Mar. 2020].

3. SOCIOLOGICAL DIMENSIONS OF AN AI SYSTEM

Lack of Diversity and thereby Biasness

A large part of the societal exclusion as a result of AI happens due to the lack of diversity in hiring practices of teams working on machine learning and AI. In order to build AI systems which do not replicate the same biases found in society, the government will have to be cognizant of the demographics of engineering teams which build these systems. A primarily male team will not do justice to the design, implementation and evaluation which AI-powered technologies require.[55] Each different stage, of producing and using AI, should have women and minorities involved in the process. Design bias encodes problems at the first step of building technology. Enough emphasis must be laid on adjusting measurements which are seen as "average" or "general". Otherwise, certain categories of the population, i.e. women and minorities, will not be represented in the systems that are built. Many forms which are used for data collection often limit the number of gender options, for instance.

The standard is two genders (M/F) and some forms allow the option of 'other' (often without specifying what the 'other' gender is), or 'transgender/third gender'. This leads to underrepresentation or a complete lack of representation when databases are built, to be used for AI. For instance in 2009, the Election Commision of India (ECI) responded to demands by activists to make the electoral process inclusive of non-male and non-female genders. Transgender people had to form activist groups and advocate for an 'other' option on the election ballot form.[56] This was a significant change, because transgender people were earlier clubbed into 'male' or 'female' categories. When thinking of all the other public systems which require forms, collect data and use AI trained on that data for the benefit of citizens; it is vital to think of inclusivity from the first step - design. It is not enough to have women for token representation, the category of 'women' must be taken to include members of the LGBTQIA+ community, women across different castes, classes, from tribal communities and those with disabilities. There are areas in design which are historically left unaddressed because of predominantly male design and engineering teams, and this bias can be fixed with diverse hiring practices. Non-binary people still struggle to find representation on datasets, since there

[55] Artificial Intelligence: Open Questions About Gender Influence, http://webfoundation.org/docs/2018/06/AI-Gender.pdf, accessed Aug 24th, 2020.
[56] Securing Transgender Rights, Economic and Political Weekly, https://www.epw.in/engage/article/securing-transgender-rights-education-finance.

is no option for them to choose their identity on forms - the difference between 'other' and 'transgender' is significant, and cannot be taken to mean the same.

Importantly, The Transgender Persons (Protection of Rights) Act, 2019 and the recent Transgender Persons (Protection of Rights) Rules, 2020 have various issues in their understanding of transgender persons, and are being criticized for resorting to male and female binaries. This means that though transgender people can 'self-determine' their gender identity - it *must* be either male or female. Such a decision leaves out those who are non-binary or gender fluid. In terms of gender reassignment surgery, the involvement of bureaucracy by way of a district screening committee is problematic and will violate the rights of transgender and transsexual people. The principle of self-determination of gender identity and the concept of district screening committees are not in harmonization with each other. In terms of datasets, there is a long way to go for India when it comes to genders on public forms. To grapple with this issue, the Government of Netherlands entirely erased the option to choose gender markers from national identification documents in July 2020.[57]

Recommendations:
Interestingly, the problem of bias in AI can be solved using AI. That is to say, using AI in hiring can help eradicate the inherent (possibly unconscious) bias and variability (noise) that comes as a result of human decision-making in hiring practices. Studies show that hiring practices reinforce stereotypes of choosing people who are like us - whether they are from similar cultures, educational backgrounds - or gender. To break out of that bias, it is important to:

(a) *emphasize the need for affirmative action when it comes to systems design hiring for engineers - and actively recruit a certain percentage of women and minorities;*
(b) *weed out the bias which exists while hiring from the general pool of applicants - by anonymizing the profiles of candidates in terms of removing gender identifiers, limiting the skill sets to only those that are required for the job (and not inadvertently including skills which might be typically masculine) and locating non-traditional applicants.[58];*
(c) *Revisit the problematic gender reassignment provisions in the Transgender Persons Act to build a more inclusive society - which will reflect in public forms, which then build datasets about the Indian population.*

[57] Examining the Right to Self Perceived Gender Identity, Livemint, https://www.livemint.com/opinion/online-views/examining-the-right-to-self-perceived-gender-identity-11597848341020.html, accessed Aug 24th, 2020.
[58] Can AI solve the Diversity Problem in Tech Industry, https://law.stanford.edu/wp-content/uploads/2019/08/Houser_20190830_test.pdf, accessed Aug 24th, 2020.

4. DISTINGUISHING DIRECT AND INDIRECT IMPACT & DISSECTING THE BLACK BOX PHENOMENON

4.1 Right To Explainability

Slide 5 of the present draft categorises challenges of the AI system on the nature of the impact i.e. direct impact (due to citizens being subject to a specific AI system) and indirect impact (due to overall deployment of AI solutions in society). However, in today's age it is nearly impossible to tell the exact impact of an AI system due to its opaqueness. We need a much clearer understanding of the normative background, the action proposed, viable alternatives, repercussions of an outcome, etc.. Thus, the distinguishing factor should not be the nature of the impact but the intention behind the attack i.e intended consequences and unintended/unanticipated consequences - a phrase popularised by American Sociologist Robert K. Merton.[59] These systems should be considered as sociotechnical assemblages made up of different institutions, personalities, locations, motives and moments. The data feeded into them is dependent on models, frameworks extracted from the existing social order and is thus not free from bias or objectivity. Thus, it is difficult for anyone to explain the AI system - issue of explainability as referred in Frank Pasquale's book: *'The Black Box Society'*[60]. This ubiquity, complexity and opaqueness leads to unintended consequences to both citizens and the society in general.

To enhance users rights, 'a right to an explanation' should be discussed and then embodied within the Personal Data Protection Bill, 2019. Some might think that right to access under Clause 17 of the Bill, 2019 subsumes right to explanation as well, however it only allows the data principal to gather information regarding personal data, a brief summary of nature of processing activities carried out by data fiduciary. Although, the nature of processing would be explained but there is a further need to explain particular decisions made by the algorithms. This right is further important for the data principal to exercise other rights - right to challenge the decision, right to obtain human intervention while decision is being made - and cannot be limited to mere '*brief summary*' of the processing activity. This right draws its basis from the National Strategy for Artificial Intelligence which draws the concept of Explainable AI systems from Pentagon's Defense Advanced Research Projects Agency (DARPA).

[59] Merton, Robert K. "The unanticipated consequences of purposive social action." *American sociological review* 1, no. 6 (1936): 894-904.
[60] Frank Pasquale, The Black Box Society (Harvard University Press, 2015)

The present draft on Slide mentions *'Fair Credit Reporting Act'* as a notable legal framework to be taken as a guidance document in future. However, the said Act only provides a reason why a particular individual was not granted a loan, rather than explaining the concerned individual what caused it. So, not only the reason for achieving a decision but also provide a counterfactual explanation[61] or reasonable recourse[62]. Though, Explainable AI systems are need of the hour but more importantly, a national common standard on explainability and transparency of AI systems (encompassing those deployed in each sector) should be framed. The framework around it is provided by us in sections 4.2 - section 4.6.

Recommendations:
(a) *Right to explanation to be added in the PDP Bill, 2019 either as a separate right or as an addition to Clause 17 of the present Bill;*
(b) *If added as a part of Clause 17, then the word 'brief summary' should be deleted and replaced with 'detailed explanation'.*

4.2 Transparency Framework

Niti Aayog going ahead can establish principles to locate Transparency to overcome the black box phenomenon as specified in **Slide 9**. We, thereby propose a set of factors which need to be considered to make an AI system more transparent, and fulfill the rule of explainability:

1. An AI system must be considered as a composite of human actors (designers, data creators, maintainers and operators) along with non human creators. In the words of Mike Ananny,[63] the systems are socio-technical assemblages in which the human agency is at least involved at the design stage. Thereby, the transparency and requirements should also be imposed upon the people designing the system.
2. The transparency information which the designers are bound to provide must rest on: a)the what i.e. the transparency of the outcomes like the prediction, assessment and, b) the how i.e. the transparency of the process of a particular algorithm to reach that particular outcome.[64]

[61] Wachter, Sandra, Brent Mittelstadt, and Chris Russell. "Counterfactual explanations without opening the black box: Automated decisions and the GDPR." *Harv. JL & Tech.* 31 (2017): 841.
[62] Ibid.
[63] Mike Ananny, "Toward an Ethics of Algorithms," Science, Technology & Human Values 41, no. 1 (2015): 93– 117.
[64] Shefali Patil, Ferdinand Vieider, and Philip Tetlock, "Process versus Outcome Accountability," in The Oxford Handbook of Public Accountability, ed. Mark Bovens, Robert. E. Goodin, and Thomas Schillemans (Oxford: Oxford University Press, 2014), 69–89.

3. Transparency information should have a standardized format which covers an array of information as different types of recipient will have varying needs according to the context and the goals of the system. A safety-inspector and an end user will need information relating to assess the system and specifics of an individual outcome respectively.[65]
4. Section 52 (ab) and (ac) of the Indian Copyright Act, 1957 provide the right to reverse engineer a particular computer software in order to determine the ideals and principles which underlie any elements of the software. The present section encourages the principle of openness so that any individual can legally check the veracity of algorithms and improve the vulnerability from the back end. This provision can be applied to any AI system as well, which automatically increases the transparency of it.

Recommendation:

(a) *AI system to be considered as involving both technical actors and human agency;*

(b) *Transparency information should consist of both 'the what' and 'the how';*

(c) *While maintaining transparency, all recipients of information should be taken care of;*

(d) *Availability of the Government AI in public domain;*

(e) *Reverse engineering under Copyright Act should be allowed.*

4.3 Accountability Framework

In addition to transparency, the present draft talks about Accountability and in detail. **Slide 12** of the Draft is solely dedicated to Accountability of AI decisions in terms of decisions by AI systems which are influenced by a complex network of decisions at different stages of its lifecycle and also assigning the burden of accountability. **Slide 16** of the draft goes to the length of emphasising on the need of a framework to assign accountability for AI systems. The Draft draws inspiration for the framework from the proposed Algorithmic Accountability Act, 2019, which would establish a law to reduce biased decisions and outcomes.

According to Kaminski and Malgieri[66], algorithmic impact assessment (AIA) is one of the means to establish accountability, though, not the sole component. However, the authors caution that AIA is best served when complemented by other accountability tools forming part of an overarching governance design. Though the draft talks indirectly about AIA, however other

[65] Alan F. T. Winfield and Marina Jirotka, "Ethical Governance Is Essential to Building Trust in Robotics and Artificial Intelligence Systems," Philosophical Transactions of the Royal Society A376 (2018).

[66] Kaminski, Margot E., and Gianclaudio Malgieri. "Multi-layered explanations from algorithmic impact assessments in the GDPR." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 68-79. 2020.

assessment approaches like Human Rights Impact assessment (HRIA), Privacy Impact Assessment (PIA), Ethical Impact Assessment (EIA) and Surveillance Impact Assessment (SIA), have not been taken cognizance of.

Establishing an AI Ethics Committee

Since the applications of AI can be detrimental to areas such as justice, health and the economy, there is a need to hold the creators of AI responsible. There is growing evidence from across the world that risk-assessment must be done proactively. In 1968, Philip K. Dick penned the dystopian novel *Do Androids Dream of Electric Sheep?* - which went on to inspire the movie *Blade Runner.* Though this is a fictional, post-apocalyptic scenario of artificial intelligence running wild, society must confront the possibility of citizens being adversely affected because of the use of AI in public services. This is grounded in research collected by authors Catherine D'Ignazio and Lauren Klein in *Data Feminism,* which provides many examples of women, people of colour and other minorities being improperly reflected in datasets and thereby not being included in the benefits AI has to offer[67]. Computer scientist and digital activist Joy Buolamwini started the Algorithmic Justice League as an organisation and movement - to expose the racial bias in artificial intelligence, particularly in facial recognition technology. There are plenty of similar examples to make a strong case for an ethics committee to regulate AI research and the deployment of AI.

In the United Kingdom, The Select Committee on Artificial Intelligence was appointed as far back as 2017. Its purpose was to "consider the economic, ethical and social implications of advances in artificial intelligence, and to make recommendations."[68] It periodically presents reports to the Parliament on the status of AI in the UK. It works to oversee that debate and policy-making are evidence-based and informed, convenes inter-sectoral experts to that end, and helps businesses and individuals to use AI to their benefit.[69] It is a 13 member committee at present, and has members from different political affiliations - Liberal Democrats, Labour, Conservative, crossbenchers and Bishops.

The White House in the United States of America has chartered a Select Committee on AI under the National Science and Technology Council. This is a horizontal committee which

[67] D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. Cambridge, MA: The MIT Press.

[68] Select Committee on Artificial Intelligence, https://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/role/, accessed Aug 24th, 2020.

[69]Government Response to AI, https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Government-Response2.pdf, accessed Aug 24th, 2020

works with other counterparts across the White House, to assist them with AI needs and Research and Development, and develops government relationships with industry bodies and academic experts.[70] The topmost officials in Research and Development in other committees of the Federal Government form the members of this committee. In its 2019 report '*The National Artificial Intelligence Research and Development Strategic Plan',* the committee recommended focussing on "understanding and addressing the ethical, legal, and societal implications for AI" as one of the key areas.[71]

***Recommendation*:**
A model framework by Andrew Selbst[72] should be adopted by the government and private sector as he mandates the assessment to be done at the pre-procurement stage and require the developers to
(a) *explain the various design choices;*
(b) *measure the resulting efficacy using the best available audit methods;*
(c) *evaluate the resulting disparate impact for the various systems and configurations*.

Given the potential of AI in India, the Indian parliament should have an ethics committee for AI specifically. It should have parliamentarians from the Rajya Sabha and Lok Sabha who are individuals who have worked in creating AI, who have an interest in proactively countering ethical conundrums with respect to AI before they arise and those who have a strong understanding of the social implications of AI technology. To that end, following from the UK and US model, it would be useful to –

(a) *Have members across political parties and affiliations;*
(b) *For the committee to produce yearly reports on the status of AI;*
(c) *An intersectoral committee which can assist other parliamentary committees on research around AI.*

4.4 Trust & Fairness Framework

The present draft culls out fairness in the context of different types of cognitive biases amplifying the large scale discrimination on the basis of gender, race, creed, religion, class and

[70] https://www.whitehouse.gov/ai/ai-american-innovation/
[71]The National Artificial Intelligence Development Strategic Plan, https://www.whitehouse.gov/wp-content/uploads/2019/06/National-AI-Research-and-Development-Strategic-Plan-2019-Update-June-2019.pdf, accessed Aug 24th, 2020.
[72] Selbst, Andrew D. "Disparate impact in big data policing." Ga. L. Rev. 52 (2017): 109.

caste. The draft makes note of Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA), which mandates certain provisions for outcome based explanations for adverse action and mandates for non-discrimination. Further **Slide 22** of the draft explores technical means to assess datasets for representation of fairness. A self-assessment guide has been prepared (**Slide 34 and 35**) which focuses on a) ensuring fairness goals while training of the system - '*Data Fairness'* b) evaluating if the system meets the fairness goals across deployment scenarios - '*Outcome Fairness'* c) Monitoring fairness goals over time and ensuring mechanisms to constantly improve - *'Fairness Sustainability'*

Lack of trust and fairness arises from the uncertainties around the system, which can be diminished if the outcome is known upfront. Also, trust and fairness is vital for any developing economy, like India, which is embedded in automation for welfare practices. Niti Aayog while building a framework around AI could consider the European Ethical principles for AI, presented by the AI4People group in 2018, suggested five principles for ethics of AI, which can be tied to trust and fairness and are explained here as follows:

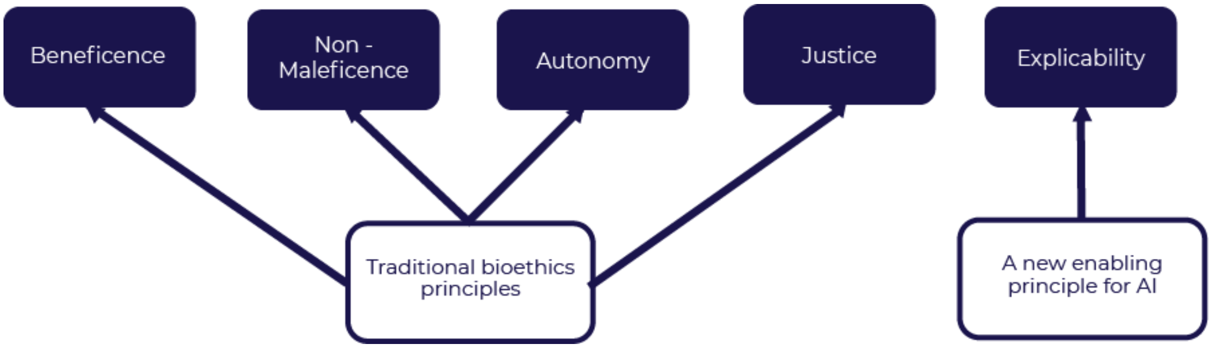
Beneficence
Non - Maleficence
Autonomy
Justice
Explicability
Traditional bioethics principles
A new enabling principle for AI

Figure 2: *An ethical framework for AI tied to trust and fairness: Floridi et.al. 'Ai4people - an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. Minds and machines'*

4.4.1 Non-Maleficence

The principle of non-maleficence states that AI should not harm people. It contains both accidental harms (overuse of AI systems) and deliberate harms (misuse of AI systems). As shown in the above sections, AI systems are inherently discriminatory in nature and therefore tend to harm individuals subjected to them. The Asilomar Principles[73] also cite the threats of an AI arms race and self-improvement of AI, as well as the need for caution around upper limits

[73] Asilomar AI Principles. (2017). Principles developed in conjunction with the 2017 Asilomar conference [Benevolent AI 2017]. Retrieved August 22, 2020 from https://futureofife.org/ai-principles

on future AI capabilities. Further, the Montreal declaration[74] argues that those developing AI "should assume their responsibility by working against the risks arising from their technological innovations". For e.g. prevention of infringements of personal privacy, listed also as one of the principles in IEEE ethics framework[75] is linked to individuals access to, and control over, how personal data is used.

4.4.2 Beneficence

The principle of beneficence states that AI shall do people good. Both the Montreal Declaration and IEEE principles use the term "well-being'', for Montreal - "*the development of AI should ultimately promote the well-being of sentient creatures*" while IEEE - "*prioritize the human well-being as an outcome in all system designs*". Asilomar principles on the other hand characterize this principle as "common good": '*AI should be developed for the common good and the benefit of humanity'*.

Ways in which an AI system may demonstrate beneficence with respect to societal problems are: a) the reduction of accidents by autonomous vehicles, b) providing support for the aging society, c) use of telemedicine via Aarogya Setu App, etc. However, this principle works in tandem with other principles and not in isolation.

4.4.3 Autonomy

This principle is the ability of individuals to make a decision for themselves, however, in the context of AI systems we willingly cede some of our decision making power. The Asilomar Principles supports the principle insofar as "humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives". Take the example of a plane cockpit where the pilot has an option to take control of the aircraft and turn off the automatic pilot system. The central point is to protect the intrinsic value of human choice and to contain the risk of delegating too much to the machines.

Thus, Floridi and others elaborate on what we need in the context of AI systems is 'meta-autonomy' where humans should always retain the power to decide which decisions to take, exercising the freedom to choose where necessary, and ceding in cases where overriding reasons may outweigh the loss of control over decision making. Herein, moral philosophy of

[74] Montreal Declaration for a Responsible Development of Artificial Intelligence. (2017). Announced at the conclusion of the Forum on the Socially Responsible Development of AI. Retrieved September 18, 2018 from https://www.montrealdeclaration-responsibleai.com/the-declaration.

[75] The IEEE Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically Aligned Design, v2. Retrieved August 22, 2020 from https://ethicsinaction.ieee.org.

Immanuel Kant: "*Kingdom of Ends*" comes into picture that every rational being should consider its moral responsibility and not mere law. Similarly, an AI system in order to be an '*ethical agent*' should exercise autonomy till it follows the other principles of the Fairness framework.

4.4.4 Justice

The concept of justice is integral to the fairness framework as it encompasses principles of accuracy, balancing the evidence and impartiality. The importance of justice is cited in the Montreal Declaration, which argues that the development of AI should promote justice and seek to eliminate all forms of discrimination. The notion of justice ensures that the use of AI creates benefits that are shared and prevents the creation of new harms, such as undermining existing social structures.

For e.g. Obermeyer and others[76], show that after an in-depth examination of AI systems (US Healthcare system) used to predict healthcare needs, they have a central issue of problem formulation. The system assumes that those individuals that will have the highest healthcare costs will be the same individuals that need the most healthcare. It is a reasonable assumption however it overlooks the fact that healthcare needs also become costly due to lack of access to transportation, child care, work related demands etc. Thus precise measures and factors needed to build a computational algorithm in any specific industry include several distortions and thereby structural inequalities, rendering failure of justice.

Thus, to ensure that AI works in a just and unbiased way we have to alter how we use machine learning and AI. The data feeded into the AI should obviate systematic biases in their AI programming and it is the responsibility of the developers.

Recommendations:

(a) *Quarterly auditing should be done by an Independent auditor of all Government AI systems or PPP projects;*

(b) *If the recommendations of the independent auditor are sent to the Public Accounts Committee of the Parliament, they should be binding upon the government;*

c) The Parliament of India should debate and discuss the Code of Ethics of AI systems and what checks and balances can be put to the Private AI systems before and after deployment. They

[76] Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366, no. 6464 (2019): 447-453.

can take cues from the Artificial Intelligence Committee of the UK though not solely rely on it.[77]

4.4.5 Explicability

Refer to 4.1 for more details on 'Explainability'.

4.5 Assessing Technical Framework

As the NITI Aayog report in June 2018 highlights, that Automated predictive systems are already in use in at least five sectors i.e. Agriculture, Education, Healthcare, Infrastructure and Mobility. Other areas not discussed in the report include, Financial, defence, law enforcement and for recruitment purposes, facial recognition technology. As Nicholas Diakopoulas argues[78], that before seeking accountability of algorithmic systems, it is necessary to ascertain the harm imposed by them, where transparency of the system comes into picture.

The Draft misses the essence of categorisation of techniques to be deployed at different stages of the AI system: Scope of the system, Design and Deployment. But, **Slide 22** of the Draft explores technical means to mitigate risks and highlights some techniques classified into Pre-Hoc Techniques and Post-Hoc techniques. The Draft also stipulates that processing of data should be in a manner which is privacy preserving, wherein techniques like federated learning, differential privacy and Homomorphic Encryption can be used. Fully Homomorphic Encryption (FHE) or Computing on the Encrypted Data is considered to be the Holy Grail of the cryptologists and along with Differential Privacy can be a strong privacy safeguard for individuals. However, on the lines of Niti Aayog Report, 2018, which does discuss the importance of anonymisation, the present draft is silent and does not go a step further to define strong anonymisation techniques.

Proper anonymisation techniques are critical to minimise risk for re-identification of these data sets. All data sets, including the initial identification data i.e., name, age, sex, profession, travel history and geolocation at time of collection and sharing of data must be anonymised. Anonymisation should be deployed to ensure a three-fold protection from: a) Singling Out - is it still possible to single out an individual, b) Linkability - is it still possible to link records relating

[77] House of Lords, Select Committee on Artificial Intelligence, "*AI in the UK: ready, willing, able*", Report of session 2017-19, Available at https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.
[78] Nicholas Diakopoulos, "Algorithmic Accountability: Journalistic Investigation of Computational Power Structures," Digital Journalism 3, no. 3 (2015): 398– 415 in *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, et al., Oxford University Press, Incorporated, 2020.

to an individual, and c) Inference - can any personal or non-personal identifiable information be inferred concerning an individual? The Government should allow independent auditors to regularly check the anonymised data sets. A subject matter group should be formed from experts in the Indian Computer Emergency Response Team and National Informatics Centre or any other relevant nodal agency, to discuss the main strengths and weaknesses of each technique. It would help to design an adequate anonymisation process in this context. Also, the technology deployed to anonymise personal data must be announced publicly on the website.

Recommendations:

(a) *Usage of the word Fully Homomorphic Systems rather than only Homomorphic should be used as the former one is more secure and recent version;*

(b) *Anonymization Techniques should be listed down as they are essential while data collection and data sharing;*

(c) *Bug Bounty programs should be a quarterly feature where Indian citizens should be allowed to plug security vulnerabilities and code improvements in any Government AI system.*

4.6 Human Rights Impact Assessment

4.6.1 Need for Human Rights Impact Assessment

With the advent of Industrial revolution 4.0 there has been a dynamic increase in data-intensive technological trends that have brought into focus the issues concerning data processing. Such a trend has compelled technology experts and human rights scholars to move beyond the traditional data protection sphere and evaluate the effect of data collection, processing and usage on the touchstone of basic human rights of the people.[79]

Innovations in AI are being deployed in the private and public sector to advance social good, and to facilitate human welfare. For example, machine learning and the availability of data are helping doctors diagnose medical conditions faster and more accurately[80]; improvements in visual recognition are helping people living with visual impairments to understand objects[81], people and text; and in road safety one study predicts that AI empowered self-driving cars can

[79] Mantelero, A. (2018). AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, *34*(4), 754-772.

[80] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

[81] Id.

help prevent up to 90% of traffic accidents, which are currently responsible for over 3,200 deaths each day.[82]

4.6.1.1 To Mitigate Social Biases

However, India being a land of diversity is abode to different communities with varying socio-economic status. Such a wide distinction in social and economic status is the cause of discrimination against and backwardness of certain communities including women, transgenders, religious minorities and people belonging to the SC/ST communities. Therefore, in light of these facts and as discussed in the section on inclusivity and non-discrimination, it is incumbent for the policy makers to take extra precaution that any new technological development is aimed at mitigating such biases.

For instance, if an AI (say, Talview) is deployed to screen CVs for a job position then it is important to ensure that the screening is solely based on merit and the algorithms are not corrupted by unwarranted biases of caste, religion or gender.

4.6.1.2 To Prevent Monopolisation of AI Technologies

As discussed in the section on AI garage the monopolisation of AI datasets in the hands of a few global players is furthering societal biases and are leading to unworthy 'best practices'. Further, it also undermines the Right to Choice, Right to Fair Opportunity and Right to Equitable Distribution of Resources of the people. Hence, a human rights impact assessment would be beneficial for ensuring an effective conflict of interest check and ensuring fair competition in the market.

In lieu of these facts human rights impact assessment of AI technologies becomes significant to help researchers gauge the potential impact of a technology on the overall living ecosystem in general and lives of the people who are most likely to be affected by that technology in particular.[83] Such an assessment would then help in designing human centric and ethically robust technologies which are steered towards the attainment of our sustainable development goals vide eliminating the biases of gender, community and region.

[82] Dignum, V. (2017). Responsible artificial intelligence: designing AI for human values, 32.
[83] Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, *361*(6404), 751-752.

4.6.2 Framework for HRIA

4.6.2.1 Determining Goals

To this end, some of the critical goals of a forward looking HRIA can include:

(a) Identifying potential risks related to the research and development (R&D) and sales of AI products and services;

(b) Respecting human rights through products, services, business activities and relationships;

(c) Informing the public debate about benefits and risks of AI and effective policy recommendations, and;

(d) Positioning the responsible use of AI as a technology respecting human rights.

4.6.2.2 Determining the underlying Rights and Regulations

The self assessment committee and expert review committee mentioned in the third and fourth prongs of this framework must test the implications of the proposed or existing AI systems based on the Principles enumerated in Part III and IV of the Constitution of India[84], Statutory Protections given under our social welfare legislations[85], and also Principles of Public International Law, including Customary International Law Principles.[86]

4.6.2.3 Self Assessment of the Existing and Proposed AI systems

As part of the HRIA legal framework, public authorities should be required to conduct a self-assessment of existing and proposed AI systems. This self-assessment should evaluate the potential impact of the AI system on human rights taking into account the nature, context, scope, and purpose of the system. Where a public authority has not yet procured or developed a proposed AI system, this assessment must be carried out prior to the acquisition and/or development of that system.

[84] Equality before Law and Equal Protection of Laws; Article 14, Right to Life and Personal Liberty; Article 21,State to secure a social order for the Promotion of welfare of the people; Article 38.

[85] The Sexual Harassment of Women at Workplace Prevention, Prohibition, and Redressal Act 2013, Rights of Persons with Disability Act, 2016, Transgender Persons (Protection of Rights Act), 2019 et. al.

[86] Universal Declaration of Human Rights; Arts. 2,3,12,18,19,20,23,25, Convention on the Elimination of All forms of Discrimination Against Women; Art. 2, Child Rights Convention; Article 2, Customary International Law, Principle 1 (The Principle of Distinction Between Civilians and Combatants), et. all.

4.6.2.4 External Review of the Existing and Proposed AI systems

The HRIAs must also include a meaningful external review of AI systems, either by an independent oversight body or an external researcher/auditor with relevant expertise, in order to help discover, measure and/or map human rights impacts and risks over time. Public bodies should consider involving the National Human Rights Commission or the relevant State Human Rights Commission for carrying out a meaningful external review.

4.6.2.5 Combating Risks Exposed in Self Assessment or External Review

In circumstances where the self-assessment or external review discloses that the AI system poses a real risk of violating human rights, the HRIA must set out the measures, safeguards, and mechanisms envisaged for preventing or mitigating that risk.

In circumstances where such a risk has been identified in relation to an AI system that has already been deployed by a public authority, its use should be immediately suspended until the above mentioned measures, safeguards and mechanisms have been adopted. Where it is not possible to meaningfully mitigate the identified risks, the AI system should not be deployed or otherwise used by any public authority.

Where the self-assessment or external review discloses a violation of human rights, the public authority must act immediately to address and remedy the violation and adopt measures to prevent or mitigate the risk of such a violation occurring again.

4.6.2.6 Public Disclosure of the Self Assessment or External Review Report

The HRIAs, including research findings or conclusions from the self assessment and external review process, must be made available to the public in an easily accessible and machine-readable format.

4.6.2.7 Allowing 'Limited Exemption' From Public Disclosure of the HRIA Report on Grounds of National Security

Only in restricted conditions the mandate of public disclosure of the HRIA report should be waived off when it appears that disclosure shall cause unprecedented national security violations.

4.6.2.8 Conducting HRIA on a Regular Basis

HRIAs should be conducted on a regular basis, and not only at the point where public authorities acquire and/or develop AI systems. It should, at the very least, be undertaken at each new phase of the AI system lifecycle and at similarly significant milestones.

Recommendations:

(a) *Ensuring a periodic HRIA of all the proposed and existing AI technologies;*

(b) *The HRIA must entail a Discrimination Assessment, Monopolisation Assessment and Conflict of Interest Assessment within its ambit;*

(c) *The HRIA should consist of two levels of assessment, the first being an internal assessment by the organisation's Ethics committee followed by an external review by independent stakeholders and human rights experts;*

(d) *Mandatory Public Release of the HRIA report.*

5. CONCLUSION

The foundational **Working Document: Towards Responsible #AIforAll** by NITI Aayog has streamlined the discussion started by **National Strategy by Artificial Intelligence** on AI in India by bringing in the focus on ethical and responsible AI. The document has been successful in identifying core issues with the deployed use cases of Artificial Narrow Intelligence while taking into account the constitutional values and the international standards. The principles of inclusivity and non-discrimination, equality, transparency, accountability, privacy and security, reliability and reinforcement of positive human values, laid down by the document are holistic and address the issues raised.

Through this report, The DIalogue has outlined the constitutional morality that should be upholded by the overarching framework of Responsible AI that NITI Aayog is planning to draft while analysing the potential of the impact of AI technologies on human rights and the society. The team has given several recommendations to address the digital divide by increasing the focus on infrastructural development for research and innovation of AI use cases in the underdeveloped states, increasing representation in the data sets collected for modelling and training algorithms, increasing the frequency of independent audits for impact assessment and ensuring transparency of the goals behind development of AI use cases while ensuring environmental safeguards are in place. This document has also suggested clause 17 in the PDP Bill, 2019 to include Right to Explanation of the use of personal data and explained the need to implement a Data Governance Framework for both Personal and Non-personal data.

In this document, the team has briefly outlined frameworks such as Algorithm Impact Assessment, Human Rights Impact Assessment and Trust and Fairness Framework which can aid NITI Aayog in building a comprehensive framework that assists India in achieving #AIforAll. Once this overarching framework is in place, it should then be followed by sector specific regulations to account for varying degrees of implementation in the sector, the stakeholders involved and so on along with adapting the existing laws dealing with labour and intellectual property rights, to the deployment of AI technologies.

About The Dialogue

The Dialogue is an emerging public-policy think-tank with a vision to drive a progressive narrative in India's policy discourse. Founded in 2017, we believe in facilitating well-researched policy debates at various levels to help develop a more informed citizenry, on areas around technology and development issues.

Our aim is to enable a more coherent policy discourse in India backed by evidence and layered with the passion to transform India's growth, to help inform on public-policies, analyse the impact of governance and subsequently, develop robust solutions to tackle our challenges and capitalise on our opportunities. To achieve our objectives, we deploy a multi-stakeholder approach and work with Government, academia, civil-society, industry and other important stakeholders.