IT for Change’s Submission to the Ministry of Electronics and Information Technology Sub-Committee Report on AI Governance Guidelines Development

IT for Change

February 2025

IT for Change's Submission to the Ministry of Electronics and Information Technology Sub-Committee Report on AI Governance Guidelines Development

IT for Change[1]

February 2025

We welcome the Ministry of Electronics and Information Technology ("MeitY") sub-committee's Report on AI Governance Guidelines Development[2] and appreciate the recognition of a lifecycle, ecosystem, and whole-of-government approach to evolving an actionable AI governance framework, keeping in mind AI's potential, challenges as well as the need to minimise risks and harms. Our detailed feedback on specific paragraphs is as under.

II. A. AI Governance Principles

While we are encouraged by the Report's emphasis on the development, deployment, and use of AI systems in ways that do not discriminate or perpetuate biases (Principle 5), a specific inclusion for adopting an integrated and indivisible human rights approach is required. This is in line with the OECD AI Principles[3] calling upon AI actors to respect human rights and internationally recognised labour rights, and NITI Aayog's Report *"Towards Responsible AI for All"*.[4]

III.A. The need to enable effective compliance and enforcement of existing laws.

1. Deepfakes/fakes/malicious content

It is submitted that the existing legal safeguards/instruments[5] may not be adequate for the

[1] To know more about IT for Change visit https://itforchange.net/

[2] Report on AI Governance Guidelines Development https://indiaai.s3.ap-south-1.amazonaws.com/docs/subcommittee-report-dec26.pdf

[3] Respect for the rule of law, human rights and democratic values, including fairness and privacy (Principle 1.2) OECD AI Principles. https://oecd.ai/en/dashboards/ai-principles/P6

[4] *"With the changing job landscape recognise and safeguard the interests of citizens under new job roles, such as gig workers"* and *"build human capacity to adapt to the changing landscape through the introduction of incentives and programs for lifelong learning and relevant reforms to education and skilling".* NITI Aayog (2021). Towards Responsible AI for All. *Approach Document for India* https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf

[5] *"The existing IT Rules only address the instances wherein the deepfake content has already been uploaded and the resultant harm has been suffered; instead, the regulatory bodies are required to put more emphasis on preventive measures, for instance, making users aware that they are looking at a morphed image."* Vig,S.(2023). Regulating Deepfakes: An Indian perspective. *Journal of Strategic Security* 17,

regulation of deepfakes.[6] For instance, Section 66E, Information Technology Act, 2000 ("IT Act") pertains to only the capture, publication, and transmission of images of private areas. Similarly, Sections 67A and 67B of the IT Act pertain to punishment for materials containing sexually explicit acts. unfortunately, deepfakes that may not fall under these legal categories are left out. The Report has further referenced legal provisions pertaining to cheating [Section 66D, IT Act, Sections 419, 463, and 465, Indian Penal Code, 1860[7] ("IPC")] as a means to protect against deepfakes. Cheating, under the IPC (Section 415) and similarly under Bharatiya Nyaya (Second) Sanhita[8] (Section 318) involves inducement to deliver any property or inducement to do or omit to do anything which the deceived person would not if he were not deceived. In many cases, the viewers of deepfakes will not be induced in the manner detailed—thus reflecting the inadequacy of the laws. Further, in all instances, even if the use of AI is not malicious, the general public has the right to know the provenance of a piece of content—whether it was AI or user-generated.[9]

As may be expected, citizens are highly concerned over the impact of deepfakes, as evidenced by the Public Interest Litigation in the case of *Chaitanya Rohilla vs. Union of India W.P.(C) 15596/2023 & CM APPLs.62399-62400/2023*[10] (*Chaitanya Rohilla*) wherein the petitioner has sought directions to identify websites giving access to deepfake AI, blocking these and ensuring accountability. The Hon'ble Delhi High Court's order dated 21 November 2024[11] in this case reflects that the Union of India/MeitY are working on projects to detect fake speech, detect and flag AI-synthesised fake images/ videos, create a web portal facilitating crowdsourcing of deepfake media, etc. In our view, such tools along with a robust regulatory framework are of the essence.

The Court, in the above-cited order, has also directed that the Committee relating to the issue of deepfakes "shall also consider the regulations as well as statutory framework in foreign countries like the European Union."

no. 3 (2024) : 70-93.
https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=2245&context=jss#:~:text=The%20existing%20IT%20Rules%20only,looking%20at%20a%20morphed%20image.

[6] Union Minister Ashwini Vaishnaw stated the government will introduce new regulations (with actionable items on four pillars—detection, prevention, reporting mechanisms and awareness) to tackle the "new threat to the society" i.e. deepfakes. DHNS (2023). New regulation to check deepfakes coming up: Govt. *Deccan Herald.* https://www.deccanherald.com/india/new-regulation-to-check-deepfakes-coming-up-govt-2783026

[7] Indian Penal Code. https://www.indiacode.nic.in/repealedfileopen?rfilename=A1860-45.pdf

[8] The Bharatiya Nyaya Sanhita. https://www.mha.gov.in/sites/default/files/2024-04/250883_english_01042024.pdf

[9] Misinformation and disinformation is the top risk identified for India by the World Economic Forum Executive Opinion Survey. World Economic Forum. (2024). The Global Risk Report. *WEF*. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

[10] https://dhcappl.nic.in/dhcorderportal/GetOrder.do?ID=mmh/2024/100018561732281370348_21590_155962023.pdfhttps://dhcappl.nic_.in/dhcorderportal/GetOrder.do?ID=mmh/2023/100018561701785455000_27500_155962023.pdf

[11] https://dhcappl.nic.in/dhcorderportal/GetOrder.do?ID=mmh/2024/100018561732281370348_21590_155962023.pdf

Recommendations:

- Existing penal laws should be suitably amended and new penal provisions should be enacted to address the legal lacunae in addressing deepfake issues. The laws should address the rights violations suffered by a person whose deepfake video or image is created and disseminated, such as his/her right to privacy, bodily integrity, dignity, reputation, etc.

- Disclosure requirements similar to those under Article 50(4) of the EU AI Act (which pertains to Transparency Obligations for Providers and Deployers of Certain AI Systems, and provides that the deployers of AI systems that generate deepfakes shall disclose that the content has been artificially generated/ manipulated) should be incorporated in the Indian legal landscape as well.

- Digital platforms such as social media, search engines, media hosting platforms, etc., should be required to institute mechanisms to proactively detect illegal and harmful content that is generated using AI tools and remove it or label it as appropriate. Platforms should also be required to adopt technical measures such as triggering an internal viral circuit breaker to prevent the algorithmic amplification of AI-generated unlawful or harmful content. Further, platforms should engage human reviewers and collaborate with fact-checkers to determine the authenticity of a piece of content. Safe harbor protection should not be available to platforms if there is a systematic or deliberate failure and gross negligence on their part by continuing to host AI-generated unlawful or harmful content.

2. Cyber security

With respect to existing laws governing cybersecurity, the Report has pointed out that the Digital Personal Data Protection Act, 2023 (“DPDPA”) requires data fiduciaries to protect personal data by putting in place appropriate security safeguards. With a view to operationalise the provisions of the DPDPA, the draft Digital Personal Data Protection Rules, 2025 (“Rules”) have been released for feedback/ comments. Rule 6, in particular, provides for the reasonable security safeguards that data fiduciaries shall undertake and these include “appropriate data security measures”, “appropriate technical and organisational measures to ensure effective observance of security safeguards." Here we submit that the lack of guidance[12] on what constitutes appropriate measures, and what involves effective observance of security safeguards, provides scope for discretion as well as dilution of data rights.

Recommendation:

- To ensure meaningful personal data protection, security safeguards must be robust and not left to interpretation. For instance, Article 25, General Data Protection Regulation (GDPR) provides for data protection by design and by default—this includes the implementation of appropriate technical and organisational measures for ensuring that, by default, only personal data that is necessary for each specific purpose of the processing is processed.

3. Intellectual property rights

Intellectual property rights (IPR) violations resulting from AI are a pressing issue to be addressed and our comments on the Report's engagement with the copyright issue are detailed below. However, we would like to highlight that it is equally important to ensure that IPR, particularly trade secret claims, are not misused by AI developers and deployers to prevent scrutiny of their models and evade liability. There is a disturbing trend of digital transnational corporations enclosing data undergirding AI systems using trade secrets, stifling genuine innovation, and making AI systems non-transparent and unexplainable.[13] For instance, in the US, trade secret claims in the information-feeding recidivism algorithms have been used to deny requests by incarcerated individuals to understand why they were given a particular rating.[14]

In a similar case, the European Court of Justice stated that companies cannot argue non-disclosure of their algorithms because of IPR or trade secrets considerations to explain AI systems within the scope of Article 22 of GDPR, except for very few considerations identified by the court, such as

[12] *"...the government must suggest technical frameworks and benchmark some international standards for technical measures like encryption, obfuscation, masking, and using virtual tokens mapped to personal information. Similarly, the government could provide some direction towards constituting access control measures and log retention. In addition to technical measures, it would be essential to benchmark some of the organisational measures…"* Shekar, K. & Sharma, V. Preliminary Analysis: Draft Digital Personal Data Protection Rules, 2025. (2025). *The Dialogue.* https://thedialogue.co/wp-content/uploads/2025/01/The-Dialogues-Preliminary-Analysis-Draft-Digital-Personal-Data-Protection-Rules-2025-Published-on-January-6-2025.pdf

[13] Kilic, B. (2024).In Uncharted Waters: Trade Secrets Law in the AI Era. *CIGI*. https://www.cigionline.org/static/documents/no.295.pdf

[14] Moore, T. R.(2017). Trade Secrets & Algorithms as Barriers to Social Justice. *Centre for Democracy and Technology.* https://cdt.org/wp-content/uploads/2017/08/2017-07-31-Trade-Secret-Algorithms-as-Barriers-to-Social-Justice.pdf

national security and criminal matters.[15] To incentivise innovation and ensure transparency and explainability of AI, India must make relevant changes to its legal frameworks related to IP and data protection to ensure that IPR, in particular, trade secrets protections, are not used to restrict data accessibility and transparency of AI systems.[16]

a. Training models on copyrighted data and liability in case of infringement

1. Questions arising on account of the intersection of AI and copyright need to be urgently addressed as the legal lacunae on issues such as whether generative AI's use of copyrighted data in order to generate responses would amount to infringement or whether the use of such data qualifies as fair use, is leading to widespread litigation. For instance, in the case of *ANI Media Pvt Ltd vs. Open AI Inc and Anr. CS(COMM) 1028/2024*[17], the Hon'ble Delhi High Court has been approached by Asian News International ("ANI") seeking an injunction restraining the defendants from storing, publishing, reproducing, or in any manner using, including through Chat GPT, ANI's copyrighted works. The questions posed by the Court[18] reflect the existing gaps in the law, and as such, there is a need to interpret and clarify the scope of copyright holders' rights in the era of generative AI.

Recommendation:

- There is an onus upon the legislature to clarify/address the legal lacunae. As stated in the Report, the list of activities for which copyrighted data can be used without permission is already provided under Section 52, Copyright Act, 1957.[19] Some scholars have argued for a fair learning principle[20] to

[15] Stankovich, M .(2024). Global Toolkit on AI and the Rule of Law for the Judiciary. *UNESCO.* https://unesdoc.unesco.org/ark:/48223/pf0000387331

[16] Gurumurthy, A. et al. (2024). Private Algorithms and Public Interest : Overhauling the Trade Secrets Regime for Equitable AI Futures, *T20 Brasil*. https://www.t20brasil.org/media/documentos/arquivos/TF05_ST_05_Private_Algorithms66cf69c5d2971.pdf

[17] https://dhcappl.nic.in/dhcorderportal/GetOrder.do?ID=abl/2024/76073424173251132546 0_14045_10282024.pdf

[18] The Court, in its order dated 19.11.2024, available at https://dhcappl.nic.in/dhcorderportal/GetOrder.do?ID=abl/2024/760734241732511325460_14045_10282024.pdf, has stated that some of key issues that warrant consideration include:
"I. Whether the storage by the defendants of plaintiff's data (which is in the nature of news and is claimed to be protected under the Copyright Act, 1957) for training its software i.e., ChatGPT, would amount to infringement of plaintiff's copyright.
II. Whether the use by the defendants of plaintiff's copyrighted data in order to generate responses for its users, would amount to infringement of the plaintiff's copyright.
III. Whether the defendants' use of plaintiff's copyrighted data qualifies as 'fair use' in terms of Section 52 of the Copyright Act, 1957.
IV. Whether the Courts in India have jurisdiction to entertain the present lawsuit considering that the servers of the defendants are located in the United States of America."

[19] https://www.indiacode.nic.in/show-data?actid=AC_CEN_9_30_00006_195714_1517807321712&sectionId=14572&sectionno=52&orderno=70

[20] *"If the purpose of the AI's use is not to obtain or incorporate the copyrightable elements of a work but to access, learn, and use the unprotectable parts of the work, that use should be presumptively fair…"* Lemley, M.A. & Casey, B. (2020). Fair Learning. Stanford Law

be used to assess copyright violation by AI models. The test herein is to see if the output of the AI system does not pose 'significant substitutive competition' to the authors/content creators whose work has been used to train the AI.[21] This test is suggested to balance the rights of creators and the promotion of AI innovation. Hence, to adequately respond to the issues arising out of the advent of generative AI models there may be a need to review the copyright law so as to ascertain whether the concept of fair learning should be introduced to assess copyright violation by AI and what its contours should be

2. Copyright holders may find it difficult to enforce their rights on account of opacity of AI systems and lack of publicly available training data.

Recommendation:

- The law should require the developers of AI systems to make details of the training data available to the public—this requirement is also found under Article 53(1)(d), EU AI Act[22] and such details of training data shall enable the copyright holders with the opportunity to exercise their rights.[23] Further, due diligence requirements may be imposed upon entities training on data—these can include, in addition to details of training data, whether such data includes personal data, the manner in which the data was obtained, etc.[24] Procurers of AI systems can further ask the entities to provide details of the due diligence carried out, and thus make an informed decision prior to procurement.

3. AI opens up the risk of cannibalising knowledge commons available in the public domain, including open government data. The exception under the DPDP Act for publicly available data further expounds this risk. Indigenous and traditional art and knowledge often lacking intellectual

School. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3528447

[21] Id.

[22] *Article 53: Obligations for Providers of General-Purpose AI Models*

1. Providers of general-purpose AI models shall: (d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.
https://artificialintelligenceact.eu/article/53/

[23] European Innovation Council and SMEs Executive Agency (2024). Artificial intelligence and copyright: use of generative AI tools to develop new content. European Commission https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/artificial-intelligence-and-copyright-use-generative-ai-tools-develop-new-content-2024-07-16-0_en

[24] Goldberg, S., Taylor, C. & Marshall, J. (2024) Tailoring the IP Due Diligence Process for All. https://www.carpmaels.com/tailoring-the-ip-due-diligence-process-for-ai/#

rights protection are also at risk of unauthorised use and value appropriation by AI systems.[25] These are often used to build proprietary and closed systems, but the profits accruing from these models and their applications do not go back to enrich the commons.

Recommendation:

- Legal frameworks related to IP and data protection should be revised to ensure that the social and economic value generated from work produced by AI trained on public datasets should be fairly distributed to the relevant community. Further, access to public domain and open government data should be conditional, with purpose limitations and clear sunset clauses on use. Robust institutional safeguards must be established for social sector datasets, such as health, education, and welfare, to ensure AI models uphold public service principles and protect marginalised communities.[26]

- In the case of indigenous and traditional art and knowledge, a collective licensing regime could be instituted to limit the reuse of work in violation of the cultural commons. The services of a collective management organization (CMO) can be utilised to negotiate licenses for a specific class of work for a specified category of workers, regardless of whether they are members of the CMO.[27]

- The risk of cultural appropriation and enclosure must also be a significant factor in the assessment of 'fair learning' for an AI model.[28]

4. AI-led bias and discrimination

1. Appropriate mechanisms are requisite to deal with AI entrenched biases and ensure algorithmic fairness.[29] In addition to the scenarios listed in the Report, platform/ gig workers largely do not

[25] Generative Māori AI tools are trained on publicly available data taken from the Māori community, without any access and benefit sharing arrangements, and the data is then sold back to members of the same community.
Artificial Intelligencer Researchers Association. (2024). Generative Reo Māori AI – The Good, the Bad and the Ugly.
https://www.airesearchers.nz/blog/post/137682/generative-reo-maori-ai--the-good-the-bad-and-the-ugly/

[26] Global Partnership on Artificial Intelligence. (2024). Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity.
https://gpai.ai/projects/responsible-ai/towardsrealdiversityandgenderequalityinai/towards-substantive-equality%20in-artificial-intelligence_Transformative-AI-policy-for-gender-equality-and-diversity.pdf

[27] IT for Change. (2023). IT for Change's submission to the U.S. Copyright Office on Artificial Intelligence Study.
https://www.regulations.gov/comment/COLC-2023-0006-9159

[28] Id.

[29] Computer scientists have been developing mathematical techniques to measure if AI models treat individuals from different groups in potentially discriminatory ways. This field is referred to as "algorithmic fairness". Information Commissioner's Office. (n.d).
https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/

have the protection of statutory rights, and are susceptible to discriminatory practices.[30] While Principle 6 of the Report stresses upon how AI systems should be subject to human oversight, it is not clear how this will be enforced vis-à-vis the labor law regime.

Recommendation:

- The requirement of human oversight is increasingly pertinent in sectors implementing algorithmic management. Here the reference to EU Directive 2024/2831[31] is relevant, wherein it is provided *inter alia* that digital labour platforms should undertake an impact evaluation, every two years, of automated systems on platform workers (pertaining to conditions of work, equal treatment, etc.), and representatives of platform workers should be involved in the impact evaluation process. Similarly, UNESCO's Recommendations on the Ethics of Artificial Intelligence provide that Member States should introduce impact assessment frameworks to identify the impact of AI systems on human rights, fundamental freedoms, labour rights, environment, etc.[32] Thus, we recommend there should be a requirement upon AI developers/deployers to undertake impact evaluation (including the impact of AI on human and fundamental rights).

2. Additionally, lack of quality training data may further AI-led bias and discrimination.[33]

Recommendation:

- AI developers/deployers should ensure AI is trained using quality training data sets to mitigate risk of bias and discrimination. To prevent discriminatory outputs, AI system providers must test for systemic bias and ensure the representation of diverse datasets.[34]

[30] Worker Info Exchange. (2024). Dying for data: how the gig economy public data deficit conceals £1.9 billion in wage theft, runaway carbon emissions, and a health & safety catastrophe. https://www.workerinfoexchange.org/public-data-sharing-report

[31] Directive (EU) 2024/2831 of the European Parliament and of the Council of 23 October 2024 on improving working conditions in platform work. https://eur-lex.europa.eu/eli/dir/2024/2831/oj

[32] UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. *UNESCO Digital Library.* https://unesdoc.unesco.org/ark:/48223/pf0000381137

[33] If the training data is insufficient, the algorithms may make predictions that are systematically discriminatory for groups that are unrepresented or underrepresented in the data. UN General Assembly. (2024). Contemporary forms of racism, racial discrimination, xenophobia and related intolerance. https://docs.un.org/en/A/HRC/56/68

[34]Global Partnership on Artificial Intelligence. (2024). Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity. https://gpai.ai/projects/responsible-ai/towardsrealdiversityandgenderequalityinai/towards-substantive-equality%20in-artificial-intelligence_Transformative-AI-policy-for-gender-equality-and-diversity.pdf

3. The lack of diversity and insufficient inclusion of marginalised communities in the design, development and deployment of AI systems is a critical challenge to address bias in AI systems.

Recommendation: It is important to take positive measures across the AI ecosystem to involve women and other historically marginalised groups in technical and non-technical roles to increase diversity in perspectives. Sufficient resources should also be allocated to identify and remove barriers to diverse representation.

III.B. The need for transparency and responsibility across the AI ecosystem in India

Assessment of AI systems on the basis of the context of deployment is the right approach to govern systems in highly sensitive sectors such as health, where patients may suffer from misdiagnosis due to bias, discrimination, and lack of gender diversity in datasets.[35] However, we are in disagreement with the Report's observation that systems should be assessed under existing sectoral laws before we evaluate the need for fresh or additional laws. The existing laws have failed to secure patients' health data[36], and while a sector-specific law (Digital Information Security in Healthcare Act)[37] has been proposed, there are no AI-related provisions in the proposed law. Further, the DPDPA is not yet operational, and considering the highly sensitive nature of health data it is submitted that sectoral guidelines with robust accountability frameworks are the need of the hour.

Recommendation:

- Guidelines for health, and other sensitive sectors, are requisite in India to provide key recommendations across the AI lifecycle. One such set of guidelines has been issued by the Ministry of Singapore—the Artificial Intelligence in Healthcare Guidelines[38]—whereby a set of recommendations to encourage the safe development and implementation of AI medical devices is

[35] Rai, A., Narayan, V. & Natarajan, S. (2022). Artificial Intelligence and Potential Impacts on Human Rights in India. *UNDP.* https://www.undp.org/sites/g/files/zskgke326/files/2022-07/Report_Artificial%20Intelligence%20%26%20Potential%20Impacts%20on%20Human%20Rights%20in%20India%20%282%29%20%281%29_0.pdf

[36] Cyberattacks hit nearly 60% of healthcare organizations globally in the past year: The Wire. (2023). https://thewire.in/tech/nearly-60-of-healthcare-organisations-in-india-hit-by-cyberattacks-in-past-year-report

[37] Ministry of Health and Family Welfare. (2018). Placing the draft of "Digital lnformation Security in Healthcare, Act (DISHA)" in public domain for comments/views. https://mohfw.gov.in/sites/default/files/R_4179_1521627488625_0.pdf

[38] Ministry of Health, Singapore. (2021). Artificial Intelligence in healthcare guidelines. https://isomer-user-content.by.gov.sg/3/9c0db09d-104c-48af-87c9-17e01695c67c/1-0-artificial-in-healthcare-guidelines-(aihgle)_publishedoct21.pdf

provided. The guidelines are based on the principles of fairness (non-discrimination), responsibility (imposed upon the organisation using the AI medical device), transparency (end users should be informed that they are interacting with AI), explainability (end users should know the data sets, training protocols, etc.) and patient centricity (safeguards in design, development, implementation to ensure patient centricity).

IV.1. To implement a whole-of-government approach to AI Governance, MeitY, and the Principal Scientific Adviser should establish an empowered mechanism to coordinate AI Governance.

We are in agreement with the Report's recommendation that the Inter-Ministerial AI Coordination Committee or Governance Group ("Committee"/"Group") should enable a whole of government approach to the AI ecosystem to ensure increased visibility and assessment of potential risks. The Report has also rightly recognised the need for better datasets in order to enable fairness, accountability, and transparency in the Indian context.

Recommendation:

- In the same vein, our suggestion is to include in the list on page 14 of the Report (measures that the Committee/Group shall suggest to catalyse collaboration between departments and regulators to *inter alia* harmonise efforts and initiatives around risk inventories, promote development and deployment of responsible AI applications), the creation of a repository of public algorithms to enhance algorithmic transparency[39] in the public sector as being undertaken in jurisdictions such as the UK[40] and EU[41]. The repository can include information regarding the policy problem the algorithm shall address, its target/beneficiaries, objectives, data sources, etc. Such algorithmic transparency shall further citizens' right to access information, along with the principles of

[39] "Algorithmic transparency can be understood from a capability, principle, standard, norm, right, and duty/obligation/responsibility perspective. Its importance lies mainly in determining which actors are the "beneficiaries" of Algorithmic Transparency and to what extent the instruments used satisfy or achieve its purpose."
Global Partnership on Artificial Intelligence. (2024). Algorithmic transparency in the public sector: Recommendations for Governments to Enhance the Transparency of Public Algorithms. *OECD.*
https://wp.oecd.ai/app/uploads/2024/12/16-Algorithmic-Transparency-in-the-Public-Sector-Recommendations-for-Governments-to-Enhance-the-Transparency-of-Public-Algorithms.pdf

[40] Under the UK's Algorithmic Transparency Recording Standard, the public sector provides information about the algorithmic tools used by them and their purpose.
https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub#:~:text=The%20Algorithmic%20Transparency%20Recording%20Standard,how%20algorithmic%20tools%20support%20decisions.

[41] The Public Sector Tech Watch, which is an observatory that monitors and disseminates the use of emerging technologies within the European public sector, has a dashboard of over 1000 use cases.
https://interoperable-europe.ec.europa.eu/collection/public-sector-tech-watch

accountability and explainability.[42]

IV.2. To develop a systems-level understanding of India's AI ecosystem, MeitY should establish, and administratively house, a Technical Secretariat to serve as a technical advisory body and coordination focal point for the Committee/ Group.

1. The requirement upon the Technical Secretariat ("Secretariat") to strengthen capacity across departments and regulators has been limited to pooling multi-disciplinary expertise; however, the capacity of the regulators itself needs to be strengthened to ensure due diligence is effectively carried out at the stage of AI procurement.[43]

Recommendation:

- The Secretariat can take up the task of forming a capacity-building program for policy makers that is regularly updated to ensure future-readiness.

2. The Report further provides that the Secretariat should facilitate the development of metrics (e.g., measurement standards for assessing the environmental impact of AI in India).

Recommendation:

- We urge that the development of these measurement standards shall be done in conformity with a rights-based approach, allowing for public participation and inputs.[44]

[42] Annex D—Type of information published by the repositories (organised per stage of the AI system's life cycle) - of the report has detailed information regarding the type of information that could be disclosed in public algorithm repositories.
Global Partnership on Artificial Intelligence. (2024). Algorithmic Transparency in the Public Sector: A state-of-the-art report of algorithmic transparency instruments. *OECD.* https://wp.oecd.ai/app/uploads/2024/12/14-Algorithmic-Transparency-in-the-Public-Sector-A-state-of-the-art-report-of-algorithmic-transparency-instruments.pdf

[43] Global Partnership on Artificial Intelligence. (2024). Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity. https://gpai.ai/projects/responsible-ai/towardsrealdiversityandgenderequalityinai/towards-substantive-equality%20in-artificial-intelligence_Transformative-AI-policy-for-gender-equality-and-diversity.pdf

[44] "Recognising that, in the field of the environment, improved access to information and public participation in decision-making enhance the quality and the implementation of decisions, contribute to public awareness of environmental issues, give the public the opportunity to express its concerns and enable public authorities to take due account of such concerns."
Convention On Access To Information, Public Participation In Decision-Making And Access To Justice In Environmental Matters. (1998). *United Nations Economic Commission for Europe*. https://unece.org/DAM/env/pp/documents/cep43e.pdf

- While developing measurement metrics of the impact of AI, the Technical Secretariat should:

- Evaluate existing methods for assessing AI risks and opportunities across sectors and regions, with a focus on human rights impact assessments.
- Establish common public AI standards to ensure AI technologies are safe, secure, transparent, and aligned with fundamental rights, rather than private or political interests.
- Propose and harmonise existing AI technical standards to promote fairness, collaboration, and human-centric development.

- Apart from developing metrics, an important mandate of the Secretariat should be to systematically collect and analyze evidence on AI's broader societal, including human rights, social, economic, political, and environmental effects, as well as differing impacts, including harms on specific groups, such as, for example, indigenous peoples and other marginalised communities.

IV.3. To build evidence on actual risks and to inform harm mitigation, the Technical Secretariat should establish, house, and operate an AI incident database as a repository of problems experienced in the real world that should guide responses to mitigate or avoid repeated bad outcomes.

The acknowledgement in the Report that AI incidents go beyond cybersecurity is appreciated; however, we believe a case can be made for mandatory reporting of serious AI incidents.

Recommendation:

- Inspiration can be drawn from Article 3(49)[45], EU AI Act which provides that serious AI incidents include death, serious harm to a person's health, fundamental rights infringements, etc. Article 73[46] subsequently imposes the obligation of mandatory reporting of serious incidents upon the providers of high-risk AI systems. Such a requirement in the Indian context will help ensure the interests of all the AI actors are balanced—AI developers, deployers, and end-users.

[45] Article 3: Definitions | EU Artificial Intelligence Act. (2025). https://artificialintelligenceact.eu/article/3/

[46] Article 73: Reporting of Serious Incidents | EU Artificial Intelligence Act. (n.d.). https://artificialintelligenceact.eu/article/73/#:~:text=They%20must%20do%20this%20as,report%20it%20within%20two%20days

IV.4. To enhance transparency and governance across the AI ecosystem, the Technical Secretariat should engage the industry to drive voluntary commitments on transparency across the overall AI ecosystem and on baseline commitments for high capability/widely deployed systems.

Self-regulation and voluntary commitments may drive the industry to only selectively commit to some of the elements envisioned as part of voluntary commitments, and result in a widespread lack of control, transparency, and accountability.[47]

Recommendation:

- Baseline binding commitments for all AI systems are crucial to ensure industry-wide uniformity and effective implementation of AI regulation. Critical aspects of AI governance, such as risk assessment and reporting of incidents, should be made binding and enforceable rather than left to discretion.

IV.6. Form a sub-group to work with MeitY to suggest specific measures that may be considered under the proposed legislation like Digital India Act (DIA) to strengthen and harmonise the legal framework, regulatory and technical capacity and the adjudicatory set-up for the digital industries to ensure effective grievance redressal and ease of doing business.

The success of a digital by design online dispute resolution system is predicated upon computer/digital literacy. Computer literacy in India, in a study conducted in 2020-21 is merely 24.7%.[48]

Recommendation:

- In order to ensure ease and effectiveness of grievance redressal, a hybrid mode may be considered, giving the aggrieved party the option to choose.

[47] “While self-regulation efforts fill regulatory gaps left by the public sector and provide governments with industry and technical expertise, they do so in potentially self-interested ways that lack democratic oversight, making it possible to leave negative lasting legacies.” Wong, A. (2023). Regulatory gaps and democratic oversight: On AI and self-regulation https://srinstitute.utoronto.ca/news/tech-self-regulation-democratic-oversight

[48] Shukla, V. & Dash, S.K. (2024). Computer literacy in India needs a reboot. *The Hindu* https://www.thehindu.com/opinion/lead/computer-literacy-in-india-needs-a-reboot/article68367762.ece

- Keeping in mind the complexity and opacity of AI systems, the burden of proof on aggrieved parties should not be high.[49]

[49] Ensure rights and redress for people impacted by AI systems. (2021). *European Digital Rights.* https://edri.org/wp-content/uploads/2022/05/Rights-and-Redress-AIA-Amendments-for-online.pdf