CoRE-AI™
Coalition for Responsible Evolution of AI

COMMENTS ON THE SUB-COMMITTEE’S REPORT ON AI GOVERNANCE GUIDELINES

FEBRUARY 27, 2025

Coalition for Responsible Evolution of AI (CoRE-AI) is a multi-stakeholder initiative on AI, housed at **The Dialogue**

COMMENTS ON THE SUB-COMMITTEE'S REPORT ON AI GOVERNANCE GUIDELINES

Dated: 27th February, 2025

To the Ministry of Electronics and Information Technology (MeitY)

Subject: Submission of Comments on the Sub-committee's Report on AI Governance Guidelines

Dear Sir/Madam,

On behalf of the Coalition for Responsible Evolution of AI (CoRE-AI), we are pleased to submit our comments and recommendations in response to the Ministry of Electronics and Information Technology's (MeitY) public consultation on the Sub-Committee's report on AI Governance Guidelines.

We commend MeitY's initiative in seeking public input to ensure that the governance mechanisms for Artificial Intelligence in India are robust, inclusive, and adaptive. This public consultation is a crucial step in fostering a collaborative approach to shaping the future of AI in our country.

Our submission, attached to this letter, reflects the collective insights and expertise of our members, gathered through extensive consultations, discussions, and analysis. We have carefully reviewed the Sub-Committee's report and have provided detailed comments and recommendations. We believe that our insights will contribute to the development of a comprehensive and effective AI governance framework that reflects India's unique aspirations and values.

We appreciate the opportunity to participate in this important consultation and look forward to further engagement on this critical issue.

Thank you for your consideration.

Jameela Sahiba

Sincerely,
Jameela Sahiba
Secretariat, Coalition for Responsible Evolution of AI (CoRE-AI)

Established in July 2024, **CoRE-AI (Coalition for Responsible Evolution of AI)** is a prominent multi-stakeholder initiative hosted by The Dialogue, focused on fostering the responsible and ethical development of AI technologies. By bringing together stakeholders from government, industry, academia, startups, and civil society, CoRE-AI aims to drive collaborative efforts that address the risks associated with AI while maximizing its societal benefits. The initiative seeks to guide India's AI journey, ensuring that technological advancements align with ethical standards to benefit the broader public

For more information
Visit https://core-ai.in/

Suggested Citation
Comments on the Sub-Committee's Report on AI Governance Guidelines (February 2025) CoRE-AI

Catalogue No
CAI/WC/0225/01

Publication Date
February 27, 2025

Disclaimer
The facts and information in this report may be reproduced only after giving due attribution to CoRE-AI (Coalition for Responsible Evolution of AI)

CONTENTS

Overview

On November 9, 2023, the Ministry of Electronics and Information Technology (MeitY) established a subcommittee to formulate actionable recommendations for AI governance in India. Tasked with identifying regulatory gaps and developing a comprehensive AI governance framework, the subcommittee conducted extensive deliberations before publishing its report on January 6, 2025. This report outlines key recommendations intended to shape India's AI policy landscape.

The Coalition for Responsible Evolution of AI (CoRE-AI), India's foremost multi-stakeholder initiative on AI, housed at The Dialogue, comprising over 50 members from startups, industry, academia, civil society, and independent experts, has conducted a thorough review of the report's recommendations. Based on a series of online and in-person consultations with our members, we respectfully submit the following comments and inputs regarding the recommendations and approaches presented in the report.

1. AI Governance Principles and their Operationalisation

Report Recommendation: *The report proposes a list of* ***eight AI Governance Principles*** *including, Transparency, Accountability, Safety, Privacy, Fairness, Human-Centered Values, Inclusive Innovation, and Digital-by-Design Governance.*

Our Suggestion: While the principles are comprehensively drafted, however, it could be of value to also explore the addition of principles of explainability and social and environmental sustainability/well-being. Ensuring AI systems are explainable enhances transparency, and incorporating environmental well-being as a principle[1] acknowledges the resource-intensive nature of AI systems, particularly in terms of energy consumption and carbon footprint. It encourages the development of sustainable AI practices that align with broader climate goals, and there is a need to use green AI solutions to minimise carbon footprints.

Further, to ensure the principles practical applicability across different AI ecosystems, the framework requires further granularity, particularly in implementation strategy and contextual differentiation.

1.1 Segmented Governance Principles for B2C, B2B, and B2G Applications

While the principles are well-defined, their implementation will vary significantly across different AI use cases—Business-to-Consumer (B2C), Business-to-Business (B2B), and Business-to-Government (B2G). Each of these domains has distinct governance priorities and risk considerations. To enhance operational clarity, we recommend a differentiated approach that tailors governance principles to

[1] OECD. (n.d.). *AI principles: Accountability.* OECD.AI. https://oecd.ai/en/dashboards/ai-principles/P5

these three segments:

- B2C (Business-to-Consumer): Consumer-facing AI systems in high-risk sectors should make efforts towards fairness, transparency, explainability, and grievance redressal mechanisms to ensure trust and accountability. This is especially critical where algorithmic decisions directly affect individuals.
- B2B (Business-to-Business): AI solutions deployed in B2B settings should prioritise clear liability frameworks, accountability through contractual obligations, third-party audits, or supply chain responsibility. Given the layered AI development process, clarity on the obligations of model developers versus downstream deployers is essential.
- B2G (Business-to-Government): AI systems used in public services must adhere to public interest safeguards, transparency in decision-making, algorithmic accountability, and routine impact assessments. Ensuring that AI deployed in governance remains ethical and auditable is crucial to maintaining public trust.

This tiered governance model would allow for realistic and effective implementation of AI governance principles while addressing the unique risks and regulatory needs of each sector.

1. 2 Harmonised AI Governance Aligned with Global Standards

India's AI governance framework should be globally interoperable to facilitate cross-border digital trade and harmonisation with international AI standards. To ensure that India's AI governance framework remains on par with global jurisdictions, it would be beneficial to explore the inclusion of principles such as environmental sustainability, which find a mention in several global AI governance frameworks. This will ensure that AI governance standards enable Indian enterprises to compete in the global AI ecosystem rather than introducing region-specific requirements that may create compliance barriers.

1.3 Clarifying 'Digital-by-Design' and its Practical Implementation

The principle of "Digital by Design" needs further definition and operational clarity. While it mirrors concepts such as "Privacy by Design" from data protection frameworks, its practical implementation remains ambiguous. Without proper clarity, it can still feel like an imposition on businesses if they are too prescriptive, limiting flexibility and increasing compliance costs.

1.4 Clarity in Principles and Balancing Competing Considerations

While principles like transparency are crucial, they must be balanced against other critical considerations, including privacy and security concerns, protection of commercially sensitive or proprietary information, alignment with evolving data protection regulations, etc.

1.5 Shared Responsibility Framework (SRF) for AI Lifecycle Governance

MeitY's emphasis on an ecosystem approach is a welcome step, ensuring that AI actors remain responsible throughout the AI lifecycle. However, we suggest incorporating a Shared Responsibility Framework (SRF) to precisely delineate obligations between AI developers and AI deployers.

- AI Developers (who build foundational AI models) should be responsible for ensuring technical robustness, and providing model documentation, but should not be held accountable for all downstream use cases beyond their control.
- AI Deployers (who customise and integrate AI models into specific applications) should look to ensure responsible deployment, compliance with sectoral regulations, and mitigating application-specific risks.
- Sectoral Regulators should oversee AI deployment in sensitive areas like healthcare, finance, and public services to ensure governance principles are upheld.
- Cross collaboration and faster resolution across all three layers of the AI Lifecycle.

By adopting this structured approach, AI governance will be both implementable and adaptable to real-world applications, ensuring an effective and accountable AI ecosystem.

2. Leveraging Technology for Governance

Report Recommendation: *The guidelines advocate for a* ***techno-legal approach*** *to AI governance, integrating regulatory frameworks with technological oversight mechanisms.*

Our Suggestion: While this approach, which includes governance technology tools alongside human oversight, can enhance compliance and accountability across AI systems; however, its successful implementation requires greater clarity on scope, flexibility, and integration within existing regulatory structures. Towards this, it is crucial to maintain a balance between regulatory compliance and innovation.

For eg., techno-legal measures to trace the use of copyrighted data in AI training needs a more nuanced consideration with the commercial sensitivities associated with training data compilation. Given that training datasets constitute proprietary assets with significant competitive value, imposing broad disclosure requirements at this early stage of AI innovation could have unintended consequences. Specifically, requiring developers to disclose details about training data could undermine trade secret protections, potentially disadvantageous to Indian AI startups. It is essential to ensure that transparency mandates do not disproportionately impact domestic AI enterprises by subjecting them to more stringent disclosure requirements than those in competing jurisdictions.

Further, mandated technological interventions often require significant system restructuring and resource allocation, which could inadvertently deter businesses from investing in advanced AI development.

To ensure a measured and effective implementation of techno-legal governance, we recommend that the government:

- Assess the necessity of techno-legal mechanisms based on clear evidence that existing enforcement mechanisms are insufficient to address specific regulatory challenges.
- Avoiding rigid enforcement provisions that could create regulatory uncertainty.
- Also ensure that the current startup/MSME business ecosystem does not get disrupted by any such new approaches.

Overly rigid mandates could introduce regulatory uncertainty and slow AI adoption. Instead, a proportionate and adaptable framework will encourage compliance while fostering innovation, enabling India's AI ecosystem to contribute meaningfully to the digital economy.

3. Compliance and Mitigation Strategies for Deepfakes and Malicious AI-Generated Content

Report Recommendation: *The Report proposes using* ***technological measures*** *for enabling effective compliance, so that malicious deepfakes are detected in time and/ or are removed before they cause serious harm.*

Our Suggestion: Regulating deepfakes and malicious AI-generated content requires a balanced approach that ensures effective compliance while avoiding undue burdens on stakeholders across the AI value chain. We concur with the Report's assessment that existing legal frameworks, including the IT Act, the Indian Penal Code (and its replacement, the Bharatiya Nyaya Sanhita, 2023), and the Protection of Children from Sexual Offences Act, 2012, provide a sufficient foundation for addressing deepfake-related concerns.

Specifically, for intermediaries, the Report highlights that provisions such as Rules 3(1)(b), (c), and 3(2)(b) of the IT Rules apply to deepfakes. These provisions require online platforms to periodically inform users about their policies, enforce compliance measures, and act on flagged content within specified timelines, such as the 24-hour requirement under Rule 3(2)(b). Additionally, we acknowledge the Sub-committee's recommendation to strengthen regulatory enforcement capabilities to ensure the effective implementation of these frameworks. Given that the fundamental nature of harm caused by deepfakes remains unchanged, we believe that the existing legal provisions, when effectively enforced, are adequate from a regulatory standpoint to mitigate the risks associated with deepfake content.

The Report further places significant emphasis on watermarking (and also discusses the possibility of assigning immutable / unique identifiers to different participants in the AI ecosystem) but overlooks several critical limitations associated with watermarking technology. Direct disclosures of watermarks increase the likelihood of them being removed or rendered ineffective through common image manipulations like cropping, screenshotting, or other basic edits. In this regard, visible watermarks (those perceptible to the human eye) are especially vulnerable to tampering, removal, or counterfeiting using standard digital tools. Invisible watermarks, while more discreet, are not entirely reliable either. The same information used to detect these watermarks can also be used to tamper or remove the same, such as by generating counterfeit watermarks (i.e., spoofing) or stripping out the watermark entirely. Creating robust invisible watermarks requires an ecosystem-wide approach to ensure they remain intact at every stage of the content lifecycle, but this is still a work-in-progress.

As regards the recommendation relating to assignment of immutable/unique identities to various participants in the AI lifecycle so that their activities can be tracked and recorded to establish liability, we believe that the operational feasibility as to how such a concept can be introduced for an AI system will need to be examined. The industry will ideally need to be consulted on the same. Our reasons for this are as follows:

- While technology-driven governance can aid enforcement, the current state of traceability tools remains inconsistent and unreliable. Existing methods, such as metadata tracking, digital watermarks, and content authentication, are often circumvented through deliberate manipulation, anonymisation tools, and adversarial attacks designed to evade detection[2].
- For instance, traceability techniques used to detect deepfake content are often easy to bypass due to the rapid advancements in deepfake generation and deliberate manipulation by malicious actors. Moreover, implementing content creator traceability for AI-generated media raises substantial privacy concerns, creating a new surface area for attacks and potential misuse. By maintaining databases that connect individuals to their AI-generated content, platforms risk exposing sensitive information, making users vulnerable to hacking, identity theft, or doxxing. Additionally, such systems could be exploited for unauthorised surveillance, compromising individual privacy rights.[3]
- If this suggestion entails additional technical customisations to be made to a model before being made available in India or introduction of a third party who will manage technology artefacts, we apprehend that it could slow down adoption of AI in India and pose security/privacy concerns.

Towards this, there is a need for a more privacy-conscious approach to mitigate these risks while ensuring accountability. Instead of rigid traceability mandates, privacy-preserving alternatives should be explored. Global initiatives such as the Coalition for Content Provenance and

[2] National Institute of Standards and Technology. (2024). *NIST report on reducing risks posed by synthetic content (NIST.AI.100-4).* U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf

[3] CCOE & DSCI. (n.d.). *Deepfake detection.* Cybersecurity Centre of Excellence (CCOE) & Data Security Council of India (DSCI). https://ccoe.dsci.in/blog/deepfake-detection

Authenticity (C2PA) offer a more secure and transparent framework by embedding verifiable provenance information within digital content. Having said that, for the C2PA standard to become a successful model of detection of AI generated content, it will need to be widely adopted across the AI value chain. In addition, end-users should also have easy access to tools to detect C2PA metadata. Further, a portfolio-based approach which encourages the adoption of best-in-class technologies and processes, and is augmented by knowledge-sharing between experts from government, industry, academia and civil society, should be explored.

To sum up, to ensure the effectiveness and adaptability of synthetic content detection measures, we recommend:

- Research and development for detection technologies should be prioritised. As part of this, MeitY can foster collaboration among industry, academia, civil society, think tanks, and other stakeholders to establish industry standards for detecting deepfakes. Since there is no universally adopted standard for identifying and labelling AI-generated content, automatic detection of AI-created or altered media across platforms remains challenging. Here, it is crucial to acknowledge that content creators of generative AI content are in the best position to identify and label synthetic content accurately.

- As regards the traceability recommendation relating to assignment of immutable/unique identities to various participants in the AI ecosystem, we request MeitY to reconsider the same for the reasons highlighted by us above. We would also like to take this opportunity to request that deepfake detection not be conflated with traceability. While traceability is vital for accountability, prioritising detection offers a more immediate and effective approach to mitigate the harm caused by harmful deepfakes. Developing and deploying reliable detection technologies will better equip individuals and organisations to manage the risks associated with synthetic media and tackle the same.
- Strengthening digital literacy and public awareness initiatives to equip users with the knowledge to identify and critically assess AI-generated content. This should specifically include AI literacy to ensure that users are better able to understand not just the risks but also the benefits of AI-powered tools (provided these are used with due care). Proactive education—through awareness campaigns, educational resources, and accessible tools, developed through collaborations with stakeholders with relevant expertise, can empower citizens to navigate the digital landscape responsibly.

4. Intellectual Property Rights

A. Input Stage

Report Recommendation:*The report states that copyright law grants the copyright holder an* ***exclusive right to store, copy etc.****, creation of datasets using copyrighted works for training foundation models, without the approval of the right holder, can lead to infringement.*

Our Suggestion:

4.1 Incorrect Assumptions made in the Report

The report incorrectly interprets the act of copying and storing data during AI training as automatic copyright infringement, unable to consider how copyright law interprets it in consonance with the right of reproduction. As per the Indian Copyright law, copyright infringement occurs when reproduction supplants the original work in the marketplace, either through direct distribution or unauthorised commercial use[4]. However, AI training does not enable users to access, distribute, or commercially exploit the original expressive works, nor does it aim to replace them. The assumption that any form of storage or duplication amounts to infringement arguably misinterprets the Indian copyright law[5].

4.2 Web Scraping and how GenAI works?

Generative AI models systematically crawl and extract publicly available data from the internet, analysing vast volumes of text, images, and other media to identify patterns, structures, and relationships that inform their learning process, without replicating or storing original works in their expressive form[6]. Web scraping, when used to extract data or learn from a work without replicating its expressive form, may arguably fall outside the traditional scope of Indian copyright protection[7]. However, this assertion hinges on the nature of the scraping activity, the type of content accessed, and whether the process encroaches upon the copyright owner's exclusive rights. As long as the intention focuses on learning from or extracting non-expressive data rather than reproducing i.e., making available for exposure, or distributing copyrighted expression, such activities would arguably avoid copyright infringement under Indian copyright law.[8]

[4] Anand and Anand. (2023, February 15). *In brief: Copyright infringement and remedies in India*. Lexology. https://www.lexology.com/library/detail.aspx?g=6526199f-85cd-4291-989d-155a7dc50272

[5] Jain, S., & Agrawal, A. (2023, October 10). *Indian copyright law and generative AI*. Saikrishna & Associates. https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai/

[6] NVIDIA. (n.d.). *Generative AI – What is it and how does it work?* NVIDIA Glossary. https://www.nvidia.com/en-in/glossary/generative-ai/

[7] Saikrishna & Associates. (2024). *Indian copyright law and generative AI – Part 5: Right to exclude access?* Lexology. https://www.lexology.com/library/detail.aspx?g=2515f6a9-2944-4893-a9ec-be9a0157d6ff

[8] Saikrishna & Associates. (2023). *Indian copyright law and generative AI*. Saikrishna & Associates. https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai/

Further, training generative AI models involves a crucial intermediary stage where datasets, including copyright-protected works, are copied and stored, typically in centralised cloud servers.[9] This storage can occur in three distinct ways: continuous storage throughout the model's lifecycle, temporary storage until the model absorbs the data, or through federated learning mechanisms that avoid centralised storage entirely. Such copies are used exclusively by model developers to extract meta-information, statistical patterns, and other underlying knowledge from the content's expression, without any human exposure to the original works[10]. GenAI as a technology functions through pattern recognition and statistical modeling rather than storing or replicating expressive works[11].

4.3 Right to Reproduction - What the Indian law protects and what it does not?

Under Indian copyright law, "reproduction" refers to the creation of copies that replicate the original work in a manner that allows human perception, enjoyment, or commercial exploitation[12]. In contrast, reproduction in the context of Generative AI is a non-expressive, computational process where data is analysed and transformed into abstract patterns, without generating verbatim or perceptible copies of the original content[13]. By disregarding the established distinction between copying for expressive reproduction versus copying for analytical learning, the report incorrectly equates technical processing with commercial misappropriation. This reasoning overlooks how copyright law permits intermediate copying when it does not interfere with the original work's market value or substitute for the protected expression[14]. The goal of GenAI is not to reproduce copyrighted content verbatim but to learn structural and stylistic patterns that enable it to generate novel outputs[15]. Understanding this nuance is critical in assessing the legal and ethical implications of AI training methodologies in the context of India's Copyright law.

[9] IBM. (2020, April 20). *Infrastructure for AI: Why storage matters*. IBM. https://www.ibm.com/think/insights/infrastructure-for-ai-why-storage-matters

[10] World Intellectual Property Organization. (2024). *Patent landscape report: Generative artificial intelligence*. https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/en/1-generative-ai-the-main-concepts.html

[11] LEXIA Avvocati. (2024, February 28). *Generative AI: The technology revolutionizing artificial intelligence*. LEXIA Avvocati. https://www.lexia.it/en/2024/02/28/generative-ai-technology-revolutionizing/

[12] Government of India. (1957). *The Copyright Act, 1957 (Including Copyright Rules)*. Copyright Office, Department for Promotion of Industry and Internal Trade. https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf

[13] NVIDIA. (n.d.). *Generative AI – What is it and how does it work?* NVIDIA Glossary. https://www.nvidia.com/en-in/glossary/generative-ai/

[14] Rao, D., & Singh, S. (2020, June 10). *Acceptable use of copyrighted material*. Singhania & Partners LLP. https://singhania.in/blog/acceptable-use-of-copyrighted-material

[15] World Intellectual Property Organization (WIPO). (2024). *Patent landscape report: Generative artificial intelligence – The main concepts*. WIPO. https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/en/1-generative-ai-the-main-concepts.html

4.4 Idea-Expression Dichotomy

The Indian copyright framework, grounded in the idea-expression dichotomy, ensures that while creative expressions are protected, underlying ideas, facts, and stylistic elements remain in the public domain[16]. This principle, reaffirmed by the Supreme Court in *R.G. Anand v. Deluxe Films (1978)*[17], is particularly relevant when assessing AI-generated content. If AI systems extract non-expressive elements without replicating protected expressions, their outputs may not constitute copyright infringement. However, in instances where AI-generated content substantially replicates original works or impacts the commercial market of creators, is where the copyright infringement concerns emanate.

4.5 Key Considerations for a Balanced Approach

4.5.1 Reproduction of copyrighted content by GenAI Models

- AI platforms should strive towards implementing measures to minimise the risk of inadvertent reproduction of copyright works, including technological measures that minimise instances of the verbatim reproduction.

4.5.2 Addressing Copyright Attribution Complexities

- Given the sheer scale of AI training datasets, assigning copyright to every contributing work is a complex challenge. Collaborative platforms pose additional challenges in determining ownership and compensation. In order to introduce remuneration mechanisms, defining eligibility and equitable distribution will be critical.

4.5.3 Providing More Control to Content Creators and Website Owners

- Platforms should empower content creators and web publishers with better control over their content's inclusion in AI training datasets. Robust opt-out mechanisms, such as improved robots.txt directives and metadata-based exclusions, should be standardised to give web administrators clear and enforceable choices regarding data usage.

4.5.4 Balancing Copyright Protections with Innovation

- The Indian copyright regime could benefit from a clearer stance on how AI-generated content aligns with existing copyright protections, ensuring that both content creators and AI developers operate within a well-defined legal structure. As technology evolves, it may be worthwhile to consider adaptations or new approaches to copyright law that effectively

[16] R.G. Anand v. M/S. Deluxe Films & Ors, (1978) 4 SCC 118 (India). Retrieved from https://indiankanoon.org/doc/1734007/

[17] R.G. Anand v. M/S. Deluxe Films & Ors., (1978). *Supreme Court of India*. Retrieved from https://indiankanoon.org/doc/1734007/

address the unique challenges posed by generative AI.

4.5.5 Strengthening Transparency and User Awareness

- While full dataset transparency may not be feasible due to proprietary considerations, platforms should enhance model transparency by providing high-level summaries of datasets used for training.
- Public awareness initiatives can help users understand AI-generated content's limitations, promoting responsible use and reducing misinformation risks.

B. Output Stage

Report Recommendation: Given the requirement of 'human authorship' for copyright protection, the eligibility and scope of granting copyright for works generated by using foundation models is untested under existing law.

Our Suggestion: Traditional copyright law is rooted in the principle of human authorship, recognising creativity as an intellectual endeavor requiring originality and personal expression[18]. However, foundation models challenge this framework by producing outputs that, while novel, may not result directly from human creative choices in the traditional sense.

A forward-looking approach should consider whether human involvement in guiding, curating, or refining AI-generated content is sufficient to meet originality standards. Courts and policymakers worldwide are still debating whether AI-assisted works should be eligible for copyright protection, and if so, what degree of human intervention is required. One possible direction is recognising AI as a tool rather than an author, ensuring that copyright protection applies only when a human exercises meaningful creative control over the final output. Towards this, tests that determine 'significant human input' beyond mere prompt-giving can be explored. This essentially means that the human creative contribution to the work should be beyond merely providing a prompt to the AI, thereby ensuring significant control over the expressive elements of the work.

Another key consideration is the economic and incentive structures underlying copyright law. Granting copyright to AI-generated works without human authorship might disrupt existing creative markets, challenge enforcement mechanisms, and create ambiguity over ownership rights. Conversely, denying protection altogether could discourage investment in AI-driven creative processes and limit their potential contributions to fields such as literature, music, and design.

A balanced approach would involve refining the definition of authorship to accommodate AI-assisted creativity while maintaining the integrity of copyright's foundational principles.

[18] Government of India. (1957). *The Copyright Act, 1957 (Including Copyright Rules).* Copyright Office, Department for Promotion of Industry and Internal Trade. Retrieved from https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf

Policymakers should explore frameworks that differentiate between fully autonomous AI outputs and works where human creators exercise substantial creative input, ensuring that copyright law remains both relevant and adaptive to technological advancements.

5. Whole-of-Government Approach to AI Governance- Inter-Ministerial AI Coordination Committee/Governance group

Report Recommendation: The report proposes the formation of an ***Inter-Ministerial AI Coordination Committee or Governance Group*** *(Committee/ Group) to bring together the various authorities and institutions that deal with AI Governance at the national level.*

Our Suggestion: While effective AI governance requires a cohesive, structured, and accountable framework that ensures coordination across multiple ministries, regulatory bodies, and state governments, Inter-ministerial committees in India have historically faced challenges related to overlapping mandates, fragmented decision-making, and limited accountability. To enhance coordination, we recommend:

5.1 Comprehensive Mapping of AI Governance Initiatives

- A systematic assessment of all AI-related committees, regulatory bodies, and working groups from the past two years should be conducted.
- This mapping exercise should analyse their mandates, outputs, and effectiveness to identify overlaps, redundancies, and gaps.

5.2 Streamlining and Optimising Governance Structures

- Based on this assessment, efforts should be made to consolidate or dissolve committees with duplicative mandates or limited progress, ensuring a more efficient and coherent AI governance ecosystem.
- The governance structure should include state governments, recognising their role in AI deployment and local policy implementation. States such as Telangana and Tamil Nadu have demonstrated leadership in AI adoption, and their insights can inform a more practical and scalable national strategy.

5.3 The Inter-ministerial Committee should aim for

- **Regulatory clarity** by reducing conflicting requirements across multiple authorities.
- **Operational efficiency**, preventing bureaucratic delays and fragmented compliance obligations.
- **Scalability and innovation**, allowing AI-driven businesses to grow within a predictable and

transparent regulatory environment.

5.4 Composition of the Committee

- To enhance effectiveness, the inter-ministerial AI coordination committee should include official and non-official members, leveraging external expertise from industry, academia, and civil society. The inclusion of state representatives will further ensure that India's AI governance framework is practical, inclusive, and scalable. Reference can be drawn from current AI governance bodies globally to eliminate redundant structures.

6. Technical Secretariat

Report Recommendation: *The Report proposes that MeitY should establish and host a **technical secretariat** that brings in officers on deputation from departments and regulators participating in the Committee/ Group and experts from academia and industry.*

Our Suggestion: While a Technical Secretariat can play a critical role in setting technical standards, developing regulatory frameworks, and providing expert guidance, however, its role must be clearly defined to avoid conflicts with existing institutions such as CERT-IN, NITI Aayog, and sectoral regulators. To ensure effective implementation:

- The Secretariat should primarily function as a standard-setting body, rather than a regulatory authority, aligning with the vision for India's AI Safety Institutes.
- It should collaborate with the Bureau of Indian Standards (BIS) and other sectoral bodies that have already initiated work on AI-related standards.
- The Secretariat must harmonise India's AI standards with international benchmarks, ensuring alignment with global best practices.
- It should also facilitate state-level engagement, operationalising the whole-of-government approach across central, state, and sectoral levels.

7. AI Incident Database

Report Recommendation: *To understand the actual incidence of AI-related risks in India, the Technical Secretariat should establish an **AI incident database** and nurture reporting to it.*

Our Suggestion: As AI systems become increasingly prevalent across sectors, systematic monitoring and reporting of AI-related incidents are crucial for enhancing safety, mitigating risks, and strengthening regulatory responses. However, it is important to recognise that existing laws already provide a broad framework for addressing cybersecurity-related AI incidents. The scope of cyber and cybersecurity incidents under current regulations is comprehensive enough to cover

AI-related harms with cybersecurity implications, and the existing reporting mechanisms are sufficient to address such concerns.

The AI-specific monitoring tool or centralised repository proposed in the Report, intended to consolidate cybersecurity and AI-related incident data, should primarily source information from public records and regulatory bodies rather than placing additional reporting obligations on industry stakeholders. This is particularly relevant as companies are already subject to extensive reporting requirements under the IT Act and the DPDP Act. That said, if an AI incident database is to be implemented, it must be clearly defined, non-duplicative, and aligned with global best practices. It should also ensure the protection of confidentiality and proprietary interests, preventing undue burdens on industry while enabling effective oversight and risk mitigation.

Towards this, we propose the following recommendations:

7.1 Defining and Scoping AI Incidents

A clear and precise definition of an "AI incident" is critical to prevent overreporting and to ensure that only material risks are recorded.

7.1.1 Criteria for AI Incident Reporting

- It is critical to adopt a definition that is specific, clear, and well-balanced. A broad definition risks compelling organisations to report a large number of minor incidents, which could overwhelm reporting systems and misdirect resources from addressing most serious harms that might uniquely arise from the use of frontier AI models in cybersecurity, chemical or biological weapons in particular. This could not only undermine the investigation of critical AI vulnerabilities but also deter innovation in the sector, particularly among startups and smaller enterprises that may lack the resources to comply with overly burdensome reporting requirements.
- Phrases like "unauthorised outcomes," "unforeseeable outcomes," and "unexpected emergent behaviour" are subjective and open to interpretations, and may be an inherent part of AI systems, given their nature. Moreover, parts of the definitions such as "privacy violations" may overlap with existing reporting requirements including those under the IT Act and DPDP Act. Further, terms like "discriminatory outcomes" in the definition introduces additional subjectivity, as these outcomes are often highly context-dependent and difficult to standardise in terms of clear criteria.
- To address these concerns, we believe that a focused and technical definition is crucial to ensure that the AI incident reporting tool prioritises meaningful and actionable incidents.

7.1.2 Alignment with International Standards

- The framework should be aligned with global AI incident monitoring mechanisms such as the OECD AI Incidents Monitor, ensuring international consistency and interoperability.
- As AI is increasingly deployed as a general-purpose technology, incident reporting should account for cross-sectoral implications, recognising that AI functions within broader technological and regulatory ecosystems.

7.2 Ensuring a Balanced and Non-Duplicative Reporting Framework

- AI incident reporting should complement, rather than duplicate, existing regulatory requirements in cybersecurity, data protection, and sectoral risk management frameworks.

7.2.1 Avoiding Redundant Compliance Obligations

- Many AI-related failures will already fall under existing cybersecurity, financial, or data protection reporting requirements. Creating an additional mandatory AI incident reporting obligation could lead to overlapping, burdensome, and conflicting compliance obligations.
- The AI incident database should serve as a knowledge repository for risk mitigation, rather than an enforcement mechanism imposing penalties or punitive measures.

7.2.2 Encouraging Voluntary Reporting, with Scope for Future Review

- AI incident reporting should be voluntary rather than mandated, allowing stakeholders to adapt to the reporting process without excessive regulatory pressure.

7.3 Confidentiality, Data Protection, and Industry Participation

To encourage industry participation, reporting protocols must ensure strict confidentiality and protection of proprietary information.

7.3.1 Safeguarding Business Interests

- Companies must be assured that reporting AI incidents will not lead to competitive disadvantages or exposure of sensitive AI models, algorithms, or proprietary datasets.
- Information submitted should be anonymised or aggregated, preventing misuse or unintended legal consequences.

7.3.2 Establishing Non-Penalisation Provisions

- Regulators must provide explicit assurances that reporting entities will not be subject to legal liability under other laws, unless the incident itself constitutes a breach of pre-existing regulations.

7.4 Governance Structure for AI Incident Management

7.4.1 Assigning Reporting Responsibility to AI Deployers

- AI deployers, rather than developers, should bear the primary responsibility for reporting incidents, as risks and harms typically arise at the deployment stage.

7.5 Ensuring a Targeted and Practical Approach

Given the vast and evolving AI landscape, an all-encompassing review of every AI use case is neither feasible nor effective. Instead, we recommend:

- Prioritising high-risk AI applications that pose significant ethical, security, or economic implications.
- Using illustrative case studies across different AI typologies to provide actionable, sector-specific recommendations rather than attempting a one-size-fits-all framework.

8. Voluntary Commitments on Transparency

Report Recommendation: *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.*

Our Suggestion: While transparency is a cornerstone of responsible AI governance, fostering trust, accountability, and ethical AI deployment, however, transparency obligations must be designed to accommodate the diversity of AI companies, ensuring they remain practical, non-restrictive, and innovation-friendly.

We recommend a flexible, sector-specific approach to voluntary transparency commitments, ensuring that companies can demonstrate responsible AI practices in ways that align with their unique operational models and competitive realities.

8.1 Ensuring Flexibility in Transparency Obligations

8.1.1 Recognising the Diversity of AI Companies

- AI companies vary significantly in size, business models, and technological applications. A one-size-fits-all transparency framework may stifle innovation and impose unnecessary burdens, particularly on startups and small enterprises.
- Companies should be empowered to adopt transparency measures that are tailored to their

specific AI systems and deployment contexts, ensuring both practicality and effectiveness.

8.1.2 Aligning Transparency with Real-World Use Cases

- Instead of prescriptive, uniform requirements, voluntary transparency commitments should allow for industry-driven, adaptive practices that reflect real-world operational complexities.
- Transparency measures should focus on demonstrating responsible AI development and deployment, rather than mandating disclosures that may be disconnected from actual risks or user concerns.

8.2 Balancing Transparency with Intellectual Property Protection

8.2.1 Safeguarding Proprietary Technologies

- Many AI systems integrate proprietary algorithms, datasets, and methodologies, making indiscriminate transparency mandates a potential risk to intellectual property.
- Any transparency commitments should be structured to ensure that companies are not required to disclose commercially sensitive or proprietary information, which could expose them to competitive disadvantages or misuse by malicious actors.

8.2.2 Designing Risk-Based, Non-Intrusive Transparency Mechanisms

Transparency obligations should be designed to ensure accountability without jeopardising innovation. This can be achieved through:

- Process-based transparency (e.g., disclosure of fairness and safety assessment methodologies rather than specific algorithms).
- Outcome-oriented transparency (e.g., impact assessments or explanations of AI decision-making processes for high-risk applications).
- Differentiated disclosure standards, ensuring that AI systems with higher societal impact (e.g., healthcare, financial services) have appropriately calibrated transparency requirements.

8.3 Industry-Led and Context-Specific Transparency Initiatives

8.3.1 Encouraging Sectoral Best Practices

- AI companies should have the autonomy to adopt context-specific transparency measures, leveraging industry-driven best practices that promote trust and accountability.
- Voluntary frameworks should be developed collaboratively, involving AI developers, deployers, regulators, and civil society, ensuring a balanced and inclusive approach.

8.3.2 Promoting Self-Regulation with Governmental Support

- Instead of top-down mandates, transparency efforts should be industry-led, with governmental support in facilitating standard-setting and best practice sharing.
- Collaborative mechanisms, such as voluntary codes of conduct, self-regulatory AI charters, etc. can encourage responsible transparency practices without imposing undue compliance burdens.

9. Legal Framework

Report Recommendation: *The report 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.*

Our Suggestion: As noted in multiple policy discussions, including this report, there is no immediate need for broad or rigid AI-specific regulations, given that the AI ecosystem is still in a formative stage, and existing laws and frameworks already cover several aspects of AI governance. Introducing overly restrictive mandates too early could inadvertently stifle innovation, limit entrepreneurial flexibility, and deter AI investments, particularly for startups and SMEs that drive much of India's AI innovation. A forward-looking regulatory approach should focus on enabling experimentation and organic sectoral evolution, ensuring that regulations remain proportional to actual risks rather than hypothetical concerns. Parallelly, the government should continue to encourage development of indigenous AI models, and support start-ups through the IndiaAI mission, along with building the nation's capacity in infrastructure, dataset availability and AI talent development to boost India's AI ecosystem.

Towards this, we recommend the following:

9.1 Leveraging Existing Regulatory Mechanisms for AI Governance

9.1.1 Strengthening the Grievance Redressal Framework

- India already has a robust grievance redressal system under the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules), which provides a structured mechanism to address concerns related to content, misinformation, and accountability of digital platforms.We believe that strengthening the existing framework would be a more effective approach to address AI-specific challenges, without duplicating efforts.

9.1.2 Regulatory Adaptation Instead of Overhaul

- Existing laws and frameworks—including those related to data protection, cybersecurity, and content regulation—already cover several aspects of AI governance.

- A sectoral approach to AI governance, rather than a broad horizontal framework, would allow for context-specific safeguards while maintaining regulatory coherence across industries.

9.2 Moving Forward: Enabling Growth While Ensuring AI Safety

9.2.1 Policy Prototyping

- India can leverage policy prototyping to test and refine AI governance mechanisms in real-world conditions, ensuring that any future regulatory interventions are evidence-based and proportional to actual risks.
- This approach would allow AI developers and policymakers to collaborate on regulatory best practices without prematurely constraining AI development.

9.2.2 Continuous Policy Iteration and Global Alignment

- AI governance should be iterative, evolving in response to technological advancements, industry feedback, and global best practices.
- India can align its AI regulatory principles with international frameworks while maintaining policy sovereignty, ensuring that regulations serve India's specific economic and societal needs.

10. Safe Harbour in the Annexure

Report Recommendation: The report suggests that in cases of AI models, the ***safe harbour*** *provision (Section 79) in the IT Act, which provides legal protection to intermediaries that host or transmit third-party content online, would not be met in many scenarios. It claims that AI systems providers or deployers cannot claim safe harbour as a default and where they do, they would need to demonstrate that they have satisfied the conditions under law.*

Our Suggestion: The interpretation that any AI involvement equates to "selection or modification" is an oversimplification that risks mischaracterising the role of modern AI systems. Algorithmic processing, such as ranking, organising, or recommending content, should not be conflated with active modification, which implies direct editorial control or content alteration.

Safe harbor provisions have historically recognised the distinction between content facilitation and content modification, ensuring that platforms are not held liable for content merely because of algorithmic organisation or automated curation. This principle should extend to AI-driven systems, maintaining the longstanding legal clarity that platforms do not forfeit safe harbor protections simply by utilising AI for content processing. A clear regulatory distinction would be necessary to prevent misinterpretations that could unduly expand platform liability, ensuring that AI's role in content moderation remains aligned with existing safe harbor protections.

Moreover (and as mentioned above), AI's value chain involves multiple stakeholders – model providers, downstream developers, deployers, and end-users – each with distinct responsibilities. Liability should be distributed across this chain, reflecting each actor's role, rather than disproportionately burdening model providers. For instance, it would be neither practical nor equitable to hold model providers solely accountable for all downstream applications of their models as they cannot predict all potential applications and thus cannot identify and mitigate every conceivable risk. Instead, AI deployers can be required to ensure that the AI models they develop/integrate are used ethically and responsibly within their specific contexts. In this regard, we believe that safe harbour benefits can be extended to model providers/developers who comply with all applicable obligations. Placing the responsibility on model providers / developers for downstream applications could make compliance impossible, as they cannot predict all potential applications and thus cannot identify and mitigate every conceivable risk.

Additionally, contractual agreements among actors within the AI value chain can help define roles, responsibilities, and liability allocation. These agreements offer a flexible and customised approach to risk management, tailored to the particular circumstances of each AI deployment. Accordingly, the ability for relevant parties to negotiate and define responsibilities contractually is recommended as opposed to defining a baseline regulatory framework for allocation of liability. End-users, in turn, must utilise AI-powered systems in compliance with the terms, conditions, and disclaimers set by system providers.

Lastly, courts worldwide are engaging with the complex legal questions surrounding AI liability, often relying on established legal principles. In common law jurisdictions, this is helping create a repository of judicial precedents on the topic. Given that India is a common law country, international judicial decisions will likely have a considerable influence on the development of AI liability law in India. Indian courts, which often look to persuasive precedents from jurisdictions with similar legal traditions, will benefit from this growing body of judicial precedents. In all, we are confident that the ongoing global legal discourse on AI liability will provide valuable insights to guide Indian courts as they navigate these emerging legal challenges.

About CoRE-AI: CoRE-AI is India's largest multi-stakeholder initiative on responsible AI evolution, housed at The Dialogue. We are currently a 50+ member alliance and have diverse representation from AI startups, academicians, civil society members, industry, and AI experts. The goal is to foster collective thinking and action to contribute to India's AI journey.

For more information, kindly contact at ***secretariat@core-ai.in.***

secretariat@core-ai.in

@Core-Ai

@CoalitionforResponsibleEvolutionofAI

core-ai.in