INDIAN GOVERNANCE AND POLICY PROJECT

Comments on the Report of the Subcommittee on AI Governance and Guidelines Development constituted by the Ministry of Electronics and Information Technology (MeitY)

Authored by
Soumya AK

With inputs from
Aayushi Gupta

TABLE OF CONTENTS

A | OVERVIEW OF THE REPORT

The Indian Governance and Policy Project ***(IGAP)*** is a premier think tank focused on driving growth, innovation, and development in India's digital landscape. Specialising in areas like Artificial Intelligence ***(AI)***, Data Protection, FinTech, and Sustainability, IGAP promotes evidence-based policymaking through interdisciplinary research. By working closely with industry bodies in the digital sector, IGAP provides valuable insights and supports informed decision-making. Core work streams include policy monitoring, knowledge dissemination, capacity development, dialogue and collaboration.

We appreciate the consultative approach taken in gathering feedback and welcome the opportunity to submit our recommendations to the Union Ministry of Electronics and Information Technology ***(MeitY)*** on the Sub-committee on AI Governance and Guidelines Development Report ***(Report).*** The Committee's insights into India's unique demographic and socio-economic landscape, along with the need for strong AI governance, align with IGAP's commitment to shaping a comprehensive framework that meets India's evolving needs while fostering the responsible advancement of AI.

Our recommendations aim to enhance governance effectiveness by ensuring regulatory clarity, encouraging innovation, and mitigating emerging risks in a balanced manner. A well-structured AI governance framework will be crucial in driving ethical and responsible AI innovation.

The Report aims to align India's AI policy with global best practices while addressing the country's unique socio-economic context. It outlines key issues, conducts a gap analysis, and offers recommendations to develop a robust AI governance framework. It emphasizes ecosystem-wide responsibility, transparency, and harm minimization as key principles for AI governance. The following sections detail our feedback on the Report.

B | PRINCIPLES AND REGULATORY APPROACHES

REPORT'S OBSERVATIONS

The Report outlines certain AI governance principles aligned with the OECD principles and proposes three possible approaches for their operationalization.

The first approach, a lifecycle-based model, regulates AI systems across their development, deployment, and diffusion stages, ensuring risk mitigation at every phase. The second, an ecosystem-based regulation model, distributes responsibilities among AI actors - including developers, deployers, and end-users - ensuring accountability across the supply chain. The third, a techno-legal governance framework, integrates regulatory oversight with technological tools such as watermarking, algorithm audits, and self-assessment mechanisms to enhance compliance and mitigate AI-related risks.

Each approach reflects a broader tension between regulation and governance, and by extension, rigidity and flexibility. This oscillation, seen throughout the Report, raises questions about the committee's direction—whether it seeks a light-touch, innovation-friendly framework or a more prescriptive, risk-driven approach.

Since the Report's introduction and the extension of the feedback deadline to February 27, 2025- global regulatory trends have evolved, with several nations shifting toward pro-innovation policies with minimal regulation.

January 2025

The United States government issued the Executive Order[1] "Removing Barriers to American Leadership in AI,"[2] emphasizing rapid AI development while reducing regulatory oversight to maintain global technological dominance.

DeepSeek R1[3] is unveiled, reflecting the global sentiment that innovation should lead while regulation follows.

February 2025

The Paris AI Action Summit[4] takes place, with both the U.S. and UK[5] refusing to sign the declaration for an open, inclusive, and ethical approach to AI development—signaling a preference for innovation over stricter regulation.

The UK renames its AI Safety Institute to the AI Security Institute, [6] shifting its focus to mitigating immediate security challenges while fostering innovation-driven growth.

This shift is also evident in India's balanced approach, which advocates both technological progress and ethical AI deployment. At the Paris AI Action Summit, the Hon'ble Indian Prime Minister,[7] who was also the Co-Chair of the Summit underscored the need for governance that supports innovation while ensuring responsible oversight, taking into account its technological capabilities and socio-economic context. India's initiative to launch its own indigenous AI model[8] and further develop its computing capacity reflects this change. In light of these geopolitical shifts, India's regulatory stance on AI may merit a reconsideration to determine the most effective approach to regulation.

While the Report outlines three possible approaches to AI governance, it does not explore how they interact with its stated emphasis on flexibility and harm mitigation. The practical considerations—how each model would function in different sectors, how responsibilities would be assigned, and how conflicts might be resolved—are left unexplored.

Instead, the Report presents these frameworks in broad terms without assessing their feasibility or alignment with India's regulatory priorities, leaving a gap in its analysis of how governance should be structured.

> ***A notable omission in the Report is the lack of discussion or attempt to define AI, an issue that carries significant regulatory implications. While it acknowledges India's technology-agnostic approach, emphasizing harms and effects over rigid classifications, it does not engage with how this choice interacts with the broader governance framework it proposes. The rationale provided—that definitions may either be too broad or fail to capture technological evolution—is valid, yet it leaves open the question of how AI systems will be identified for regulatory purposes.***

The European Union AI Act **(EU AI Act)**, for instance, adopts a broad yet functional definition[9] to maintain adaptability while ensuring regulatory clarity. By avoiding this discussion, the Report shifts the burden of interpretation onto future regulatory mechanisms, potentially creating ambiguity in scope, applicability, and risk categorization.

C | LIFECYCLE AND STAKEHOLDER ROLES

REPORT'S OBSERVATIONS

The Report effectively acknowledges that AI risks evolve across different stages of development, deployment, and diffusion, and it recognizes key stakeholders within the AI ecosystem, including data principals, data providers, AI developers, deployers, and end-users.

The Report does not clearly delineate stakeholder responsibilities or examine how these roles may shift or overlap in practice, especially in industries where AI deployment is complex and evolving. It outlines different actors but lacks a responsibility matrix that would clarify how accountability is distributed across the AI lifecycle.[10]

Moreover, the Report acknowledges the need for human oversight, it does not specify how oversight mechanisms would be adapted to different risk levels, or which stakeholders would be responsible for implementation.

The EU's approach to AI governance, as reflected in its Ethics Guidelines for Trustworthy AI,[11] discusses oversight mechanisms such as human-in-the-loop (HITL), on-the-loop (HOTL), and in-command (HIC).[12] While these models are not explicitly codified in the EU AI Act, the Act's risk-based framework mandates proportionate human oversight for high-risk AI systems, aligning with these principles in practice.

Similarly, while the Report advocates for a "digital by design" governance framework, its mention of traceable artifacts as a regulatory tool warrants further scrutiny. While such mechanisms can enhance accountability and compliance, they may disproportionately impact resource-constrained actors like SMEs and startups.

RECOMMENDATION

Further clarification of stakeholder roles and responsibilities, along with a focus on supporting resource-constrained actors like SMEs, would enhance the inclusivity and practicality of the governance framework, a view echoed at the Paris AI Summit[13] that underscores the need for AI to be inclusive, ethical, and safe.

D | GAP ANALYSIS

REPORT'S OBSERVATIONS

The Report has focused on identifying regulatory gaps, emphasizing the need for effective enforcement of existing laws in AI-related harms and ensuring regulators have access to comprehensive information on the AI ecosystem, including data, models, applications, and stakeholders.

Given AI's rapid evolution and cross-sectoral impact, the Report advocates for a whole-of-government approach to address emerging risks. It explores issues such as cybersecurity, intellectual property rights, deepfakes, AI-led bias, and transparency within the AI ecosystem.

While it offers valuable insights, it falls short of fully aligning with the goal of guiding legal and regulatory improvements.

The analysis would benefit from:

i. Greater emphasis on proactive and preventive measures.

ii. Structured frameworks for risk-based classification and traceability.

iii. Engagement with global best practices to inform domestic policy.

Vi. Specific guidance on addressing cross-sectoral and public-sector challenges.

1 Deepfakes and Malicious Content

REPORT'S OBSERVATIONS

The Report addresses the misuse of AI systems to create malicious synthetic media, such as deepfakes, by referencing existing legal frameworks like the Information Technology (IT) Act, 2000 and erstwhile the Indian Penal Code (IPC) (presently, Bharatiya Nyaya Sanhita, 2023). It underscores intermediary obligations under the IT Rules to detect and remove harmful content and suggests technological tools like watermarking and traceability to enhance compliance.

(a) Enhancing Proactive Harm Mitigation

The Report emphasizes harm minimization but the analysis focuses on post-harm measures, such as takedown mechanisms under the IT Act. This post-harm focus may be insufficient for proactive harm mitigation.

For instance, in the United States, certain measures like the proposed[14] DEEPFAKES Accountability Act,[15] and the Protect Elections from Deceptive AI Act[16] are there – however, both of these remain legislative proposals and have not yet been enacted into law. The former aims to protect national security by imposing transparency requirements on AI-generated content and the latter seeks to curb the use of AI-generated deceptive materials in political campaigns. Other countries have also recognized the need for targeted legislation; for example, Taiwan[17] has criminalized deepfake-related fraud, acknowledging the unique challenges posed by AI-manipulated media.

(b) Broadening the Analysis of Deepfake Risks

The Report also lacks a nuanced analysis of the varied risks posed by deepfakes across different contexts, such as personal reputational damage versus political destabilization. Internationally, jurisdictions like the European Union have considered implementing transparency obligations for high-risk AI systems, requiring clear labelling of synthetic content[18] to inform and protect the public. The Report assumes that existing laws like the IT Act are sufficient but does not critically assess their effectiveness or identify enforcement gaps specific to AI-generated content.

(c) Strengthening Current Legal Framework and Assigning Stakeholder Accountability Across AI Ecosystem

Additionally, while the current legal framework to address deepfakes has been supplemented by efforts from MeitY, these measures largely remain advisory and lack enforceability. In December 2023,[19] MeitY issued an advisory directing intermediaries to comply with the IT Rules, 2021, focusing on misinformation and deepfakes. It stressed user communication in preferred languages, content moderation, and prompt removal of misleading material. A March 1, 2024, advisory reinforced due diligence, requiring labelling of under-trial AI models and prior government approval before deployment.[20] The March 15, 2024, revision[21] commendably removed prior approval but retained labelling for deepfakes and user notifications of unreliable AI outputs. However, ambiguities in legal criteria, voluntary adherence, and a narrow focus on intermediaries raise enforceability and compliance concerns.

The Report also focuses narrowly on platforms while overlooking other key stakeholders—developers who build AI models, deployers who integrate them, and end users who generate and share content. Assigning liability solely to hosting platforms ignores the broader chain of accountability required to address AI-generated misinformation effectively. A comprehensive regulatory approach must consider this entire ecosystem rather than relying solely on intermediaries for enforcement.

(d) Strengthening Content Authenticity with Provenance Tools

The Report's recommendations to implement immutable identities and watermarking are commendable steps toward enhancing the authenticity and traceability of digital content. In addition to these measures, provenance tools can further strengthen content authenticity by tracking the origin and modification history of digital media.[22] ***These tools allow users to verify whether an image, video, or document has been altered, providing an additional layer of transparency. Similarly, the Coalition for Content Provenance and Authenticity (C2PA),***[23] ***an open technical standard, provides publishers, creators, and consumers the ability to trace the origin of different types of media, enhancing transparency and trust in digital content. While these measures enhance transparency and content authenticity, their effectiveness will depend on clear legal mandates, industry cooperation, and enforcement to prevent the misuse of synthetic media.***

2 Cybersecurity

REPORT'S OBSERVATIONS

The Report recognizes that existing laws, such as the IT Act and associated rules, along with agencies like CERT-IN (Indian Computer Emergency Response Team) and NCIIPC (National Critical Information Infrastructure Protection Centre) provide a foundational framework for cybersecurity and outline sector-specific guidelines from regulators like RBI and SEBI. It also acknowledges the need to enhance enforcement and compliance capabilities to address AI-specific risks.

However, the analysis underestimates the unique cybersecurity challenges introduced by AI systems. These include adversarial attacks, data poisoning, and vulnerabilities in generative AI applications, which are not adequately addressed within the existing framework.[24]

Additionally, while the Report highlights existing institutional frameworks, it does not assess whether agencies like CERT-IN and NCIIPC have the technical expertise, training, and resources necessary to handle AI-driven cyber threats. AI security[25] demands specialized threat modelling, continuous monitoring, and adversarial testing capabilities, which require AI-specific capacity building within these institutions.

The Report's call for "secure by design" systems is a positive step but lacks practical implementation strategies. Notably, frameworks such as Singapore's Guidelines on Securing AI Systems[26]—which advocate comprehensive risk assessments, threat modelling, and continuous monitoring—and US NIST's Secure Software Development Framework (SSDF)[27]—which

emphasizes secure coding practices and rigorous testing—provide valuable guidance for embedding secure practices throughout the AI lifecycle.

The Report acknowledges sectoral silos in cybersecurity but lacks cross-sectoral security baselines or AI-specific collaboration mechanisms. Given AI's industry-wide risks, a unified framework—could enhance systemic resilience.

3 Intellectual Property Rights (IPR)

The copyright section in the gap analysis addresses two primary issues: (a) the use of copyrighted material for training AI models, and (b) the copyrightability of AI-generated output.

(a) The Use of Copyrighted Material for Training AI Models

REPORT'S OBSERVATIONS

The Report acknowledges that training AI models on copyrighted material could lead to copyright infringement unless explicitly authorized by the copyright holder. It notes that Section 52(1)(a)(i) of the Indian Copyright Act, 1957, provides a narrow set of exceptions, meaning that commercial and institutional research do not qualify as exemptions, making AI training on copyrighted data potentially infringing.

However, the Report presents contradictory positions—on one hand, it assumes that AI training may not be covered under current copyright law, but later, it raises open-ended policy questions about whether AI systems should be allowed to train on copyrighted datasets without explicit approval.

The Report does not engage with key legal debates regarding whether storing copyrighted material for AI training constitutes infringement[28] under the Indian Copyright Act, 1957, or whether AI training qualifies as fair dealing under Section 52. While it acknowledges that AI training on copyrighted material could be legally contentious, it does not explore the idea-expression dichotomy,[29] the role of non-expressive data processing,[30] or the implications of case law on fair dealing. These issues are currently under judicial scrutiny in India; in November 2024, news agency Asian News International (ANI) filed a lawsuit[31] against OpenAI in the Delhi High Court.[32]

(i) Ambiguity Regarding Section 52(1)(a) and the 2012 Amendment

The Report also does not clarify whether AI training aligns with the 2012 amendment to Section 52(1)(a), which expressly permits the storage of copyrighted works in an electronic medium for research and education. This amendment was introduced to accommodate technological advancements, yet the Report does not examine whether AI training[33]—which involves tokenization of the text and thereafter learning to recognize statistical patterns using neural networks rather than merely storing or replicating content—could be considered a permitted activity under this provision. This omission creates ambiguity regarding whether AI training falls within a legal exemption or constitutes infringement.

(ii) Lack of Analysis on Fair Dealing Judicial Tests

The Report does not sufficiently engage with key judicial tests that determine whether AI training qualifies as fair dealing under Indian copyright law. One of the most critical tests is the transformative use test,[34] which assesses whether the secondary use significantly alters the original work's purpose and function.

Indian courts, such as in *Syndicate of the Press of the University of Cambridge v. B.D. Bhandari,*[35] have emphasized that transformative use is a key determinant of fair dealing . However, the Report does not examine whether AI training—which involves extracting statistical relationships rather than replicating expressive content[36]—meets this threshold.

Second, the market impact test,[37] evaluates whether AI-generated outputs could substitute or compete with the original works. Courts have ruled that fair dealing is not applicable if the secondary use adversely affects the market for the original work. The U.S. case of *Authors Guild v. Google Inc.*[38] found that full-text scanning of books for search indexing constituted fair use because it did not replace the market for the original works. Similarly, AI models store data in a machine-readable tokenized format,[39] fundamentally different from human-readable copyrighted works. The Report does not discuss whether this format serves as a market substitute.

Finally, the extent of use and justification test (*Super Cassettes v. Hamar Television (2010)),*[40] which examines whether the volume of copyrighted material used is justified by the purpose of the use.

AI models require large-scale ingestion of data for effective training, raising questions about whether the sheer quantity of stored material affects its fair dealing eligibility. Courts have ruled that extensive reproduction is justified when it serves a distinct and transformative function, such as in the case of *Rameshwari Photocopy Services,*[41] where full reproduction of course materials was deemed permissible for educational purposes.

The Report does not engage with how this precedent might apply to AI training. These gaps leave significant legal questions unresolved, underscoring the need for further analysis to determine whether AI training falls within the existing copyright exceptions. The pending case before the Delhi High Court may

provide clarity on these issues, and its outcome will likely shape the legal framework unless other developments arise in the interim.

(b) The Copyrightability of AI-generated Output

A key issue raised in the Report is the copyrightability of AI-generated works. Indian copyright law, like many global frameworks, currently requires human authorship for protection, raising uncertainties about whether AI-generated outputs can qualify for copyright and under what conditions.

The Indian Copyright Act, 1957, does not explicitly address AI-generated works. However, under Section 2(d), an "author" is defined based on the type of work, and human authorship remains a fundamental requirement. It has been suggested that AI-generated works might be classified as "computer-generated works" under Section 2(ffc), which covers works created using computers. However, Section 17 presumes authorship to be human-centric,[42] creating ambiguity regarding AI's role in authorship and ownership.

These are critical areas of concern, but the Report does not sufficiently explore domestic law or how other jurisdictions are addressing similar issues. For instance, while jurisdictions like the United States, the European Union, and the United Kingdom are grappling with similar legal uncertainties, there are notable differences in how each is addressing AI-generated works. A deeper examination of emerging policy debates, legal precedents, and potential pathways for reform in these jurisdictions provides valuable insights into how India could navigate its own regulatory approach.

Global Perspectives

*1) **United States**: The U.S. position has been that human authorship is a prerequisite[43] for copyright protection. In its recent Report, [44]the U.S. Copyright Office emphasized that works entirely generated by AI without human intervention are not eligible for copyright. However, if a work includes both human and AI-generated content, only the human contributions are potentially copyrightable. The mere act of inputting prompts into an AI system does not constitute authorship, as it lacks sufficient creative control over the output.*

*2) **European Union**: The European Union's approach to AI-generated works remains cautious,[45] with most Member States agreeing that purely AI-generated content should not receive copyright protection, as only a natural person can be considered an author under the Berne Convention. Recent discussions within the Council of the European Union suggest that while the existing framework is largely sufficient, more clarity is needed on text and data mining (TDM) exceptions, licensing mechanisms, and the role of collective management organizations (CMOs) in AI-related copyright issues.*

*3) **United Kingdom**: The UK Copyright, Designs and Patents Act, 1988, explicitly recognizes "computer-generated works" (Section 9(3)), assigning authorship to "the person by whom the arrangements necessary for the creation of the work are undertaken." A Consultation document[46] (December 2024) presented to the Parliament notes that the UK currently provides copyright protection for purely computer-generated works, but it is not clear that this protection is widely used, or that it functions properly within the broader copyright framework. The government is now seeking views on whether reforms are necessary to ensure the law adequately accounts for AI-generated content.*

4 AI Led Bias and Discrimination

REPORT'S OBSERVATIONS

The Report appropriately acknowledges the risks of AI-driven biases and the limitations of existing legal frameworks in addressing systemic discrimination. It highlights the importance of transparency and accountability mechanisms to mitigate these risks.

However, the section remains high-level and lacks specific implementation strategies to detect and mitigate bias effectively. The Report overlooks high-risk AI applications, such as law enforcement, predictive policing, and public welfare programs, where biased algorithms can exacerbate societal inequalities. AI-driven decision-making in these sectors can reinforce discriminatory patterns in areas such as criminal sentencing, resource allocation, and eligibility assessments for government benefits. The risks associated with historically biased datasets and unexplainable AI models in such contexts require a more in-depth examination.

Additionally, while the Report stresses the importance of transparency, it does not provide concrete mechanisms for bias detection and mitigation. Global best practices,[47] such as algorithmic impact assessments (AIA), fairness metrics, bias audits, and independent oversight boards, could offer actionable safeguards. For example, a US Blueprint for an AI Bill of Rights[48] proposes pre-deployment testing and independent evaluations to ensure AI fairness in critical decision-making.

The Report also remains ambiguous about accountability—it does not clarify whether developers, deployers, or end-users bear responsibility for biased outcomes, which is essential for regulatory enforcement and compliance.

RECOMMENDATION

This section would benefit from a deeper examination of high-stakes use cases and more specific guidance on operationalizing bias detection and mitigation measures within the governance framework to ensure a trust-based framework that advances AI responsibly and ethically. By identifying bias risks—especially in high-stakes scenarios—our approach will align with the global sentiment[49] that while the underlying technology could remain unregulated so as to foster innovation, its use cases need to be regulated according to their risk profiles.

E TRANSPARENCY AND RESPONSIBILITY ACROSS THE AI ECOSYSTEM

REPORT'S OBSERVATIONS

The Report emphasizes the need for traceability of data, models, systems, and actors throughout the AI lifecycle and the importance of transparency in allocating liabilities across ecosystem actors. It also acknowledges the necessity of a baseline framework to ensure responsible AI deployment.

(a) Strengthening Transparency and Traceability Mechanisms

However, the Report does not provide concrete guidance on how traceability and transparency mechanisms would be implemented, particularly for SMEs that lack the resources to comply with complex regulatory requirements. Additionally, ensuring traceability must be balanced with data protection concerns, as excessive tracking could expose confidential information or conflict with existing privacy laws.

(b) Emphasising the Need for a Risk-Based Framework

The Report references high-risk AI systems, but it does not propose a structured, risk-based classification framework to guide transparency efforts. Innovation-driven objectives in the Global South could benefit from established frameworks like the EU AI Act or classification systems—such as that in the International Scientific Report on the Safety of Advanced AI[50]—which provide a useful starting point for understanding AI-related risks and sub-risks. These frameworks adopt a technology-neutral approach by classifying AI systems based on risk levels and identifying technical methods to mitigate those risks.

RECOMMENDATION

Developing such a risk-based framework would also require regulations tailored to address distinct, context-specific risks while maintaining the flexibility to adapt to the rapid evolution of AI technologies. To strengthen its relevance, this section should provide clearer recommendations on operationalizing traceability and transparency, ensuring they are practical, scalable, and aligned with risk-based governance approaches.

F | RECOMMENDATIONS

The Report highlights the inefficiencies of a fragmented regulatory landscape and advocates for a whole-of-government approach to AI governance, a well-placed emphasis given the multi-sectoral nature of AI risks and India's socioeconomic needs.

Despite this focus, the guidance on harmonizing diverse sectoral laws remains insufficient, risking duplication, jurisdictional conflicts, and inefficiencies. Greater clarity on how existing regulatory bodies will coordinate with the AI Coordination Committee and the Technical Secretariat is necessary to ensure a cohesive governance structure that avoids regulatory overlap.

The Report emphasizes harm minimization as a foundational regulatory principle, which is a commendable approach. However, it does not clearly differentiate between addressing risks (future probabilities of harm) and mitigating harms (ongoing or potential damage). This lack of clarity can hinder the development of precise governance strategies. While activity-based regulation is presented as a starting point, the Report does not explore how India will transition toward a combination approach for AI regulation as risks evolve. The absence of a clear transition roadmap, particularly for high-risk AI applications, could hinder effective governance.

Other recommendations and suggestions from the Report have been discussed below.

---

1. Strengthening coordination mechanisms:

The implementation roadmap for coordination mechanisms remains underdeveloped. While the Report suggests measures such as inter-ministerial committees, it fails to provide specific details on their structure, functions, or processes for resolving jurisdictional overlaps.

The interplay between the proposed AI Coordination Committee, the Technical Secretariat, and sectoral regulators lacks clarity, increasing the risk of duplicative efforts. A clear delineation of roles and responsibilities is crucial to prevent inefficiencies and ensure smooth coordination across various bodies.

2. Building Technical Capacity for Regulators:

While the Report acknowledges the need for capacity building through the Technical Secretariat, it does not sufficiently address how sectoral regulators will develop in-house AI expertise.

AI governance requires not just policy coordination but also technical enforcement capabilities, which many regulators currently lack. To bridge this gap, structured capacity-building programs, AI-specific training initiatives, and dedicated AI enforcement units should be established within key regulatory agencies.

3. Tailoring Industry Commitment to Sector-Specific Needs:

The Report's recommendations for voluntary industry commitments and governance frameworks are too generic to be impactful. Commitments such as transparency Reports and red-teaming exercises require sector-specific tailoring to

address the unique challenges and risks of different domains. Without such granularity, these recommendations risk being ineffective or misaligned with sectoral needs, consequently hindering innovation.

4. Optimising Technological Solutions:

The feasibility of implementing technological solutions such as watermarking or liability chains is not adequately explored. The Report does not address how these solutions will be reconciled with existing legal frameworks or technical limitations. Provenance tools, such as Microsoft's Project Origin[51] and C2PA[52] have been proposed globally as alternative transparency-enhancing mechanisms and could be incorporated into India's regulatory approach. A phased implementation plan with pilot studies would help assess the practicality and impact of these measures.

5. Empowering SMEs through Targeted Incentives:

The recommendations also focus disproportionately on larger stakeholders, overlooking the unique challenges faced by startups and small and medium enterprises (SMEs) in navigating complex governance requirements. SMEs often lack the financial and technical resources to comply with intricate regulations, placing them at a disadvantage.

Recent Developments

Meanwhile, Bank of China unveiled its AI Industry Development Action Plan[53], pledging $137 billion to all entities across the AI industry chain over the next five years, while the US $500 billion joint venture[54] with tech companies is underway to build AI-focused infrastructure and data centres within the next four years. In contrast, India's current support of ₹10,300 crore[55] and its recent initiative to procure GPUs[56]—while commendable—may fall short[57] given the rapid pace of AI innovation and intense global competition.

Sector-specific funding models, compliance assistance programs, and AI regulatory sandboxes could be introduced to support SMEs while fostering responsible innovation.

6. Promoting Transparency and Ethical Safeguards:

Finally, the Report does not sufficiently address the need for transparency and ethical safeguards in government-deployed AI systems, particularly in high-stakes areas like law enforcement and welfare. Given the significant societal impact of such applications and the recent declaration at the Paris Summit promoting ethical and responsible AI, explicit recommendations for transparency, accountability, and ethical safeguards are essential to ensure fairness and public trust.

G | CONCLUSION

In its conclusion, the Report re-emphasises harm mitigation as the core regulatory principle. The argument that "enabling innovation is itself harm minimization" risks creating regulatory ambiguity, as it could prioritize economic growth over necessary safeguards. While activity-based regulation is a pragmatic starting point, the Report does not explore how India will transition to a combination approach as AI risks evolve. Without a clear roadmap, particularly for high-risk AI applications, effective governance could be compromised. The Report makes valuable strides in proposing collaborative governance mechanisms and industry engagement, but given its potential to shape future AI policy and regulation, greater clarity is essential in implementation strategies, risk assessment frameworks, and regulatory mechanisms. Strengthening institutional capacity, supporting SMEs, ensuring sectoral coordination, and adopting a coherent framework will be key to ensuring that India's AI governance remains adaptive, accountable, and enforceable.

ENDNOTES

1 Executive Office of the President, *'Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence'*, EO 14110 (Issued on October 23, 2023) (USA).

2 Executive Office of the President, *'Executive Order on Removing Barriers to American Leadership in Artificial Intelligence',* EO 14179 (Issued on January 23, 2025) (USA).

3 Charlotte Edmond, *What is Open-Source AI and How Could DeepSeek Change the Industry?*, World Econ. F. (February 5, 2025), https://www.weforum.org/stories/2025/02/open-source-ai-innovation-deepseek/.

4 *AI Action Summit co-chaired by France and India,* La France en Inde / France in India (February 13, 2025), https://in.ambafrance.org/AI-Action-Summit-co-chaired-by-France-and-India.

5 Zoe Kleinman and Liv McMahon, *UK and US refuse to sign international AI declaration,* BBC News (February 11, 2025), https://www.bbc.com/news/articles/c8edn0n58gwo.

6 Mackenzie Ferguson, *UK's AI Safety Institute Rebrands as AI Security Institute: A Bold New Focus on National Risks,* (Feb. 14, 2025), https://opentools.ai/news/uks-ai-safety-institute-rebrands-as-ai-security-institute-a-bold-new-focus-on-national-risks.

7 *Prime Minister co-chairs AI Action Summit in Paris,* Ministry of External Affairs Press Release (February 11, 2025), https://www.mea.gov.in/press-releases.htm?dtl/39023/Prime_Minister_cochairs_AI_Action_Summit_in_Paris_February_11_2025.

8 *With robust and high-end Common computing facility in place, India all set to launch its own safe & secure indigenous AI model at affordable cost soon: Shri Ashwini Vaishnaw,* PIB Press Release (January 30, 2025), https://pib.gov.in/PressReleasePage.aspx?PRID=2097709.

9 Regulation (EU)- 2024/1689 ("EU AI Act"), Article 3(1) defines 'AI system' as "machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."

10 Blaine Keuhnert and Rachel M. Kim, *The "Who", "What", and "How" of Responsible AI Governance: A Systematic Review and Meta-Analysis of (Actor, Stage)-Specific Tools,* (February 18, 2025), https://arxiv.org/html/2502.13294v1?utm_source=chatgpt.com.

11 European Commission, *'Ethics Guidelines for Trustworthy AI,'* Publications Office of the European Union, (Issued on April 8, 2019).

12 HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system.

13 Elysee, *Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet*, (February 11, 2025), https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet.

14 Jana Kazaz, *Regulating Deepfakes: Global Approaches to Combatting AI-Driven Manipulation,* GlobSec, (December 10, 2024), https://www.globsec.org/sites/default/files/2024-12/Regulating%20Deepfakes%20-%20Global%20Approaches%20to%20Combatting%20AI-Driven%20Manipulation%20policy%20paper%20ver4%20web.pdf.

3 Charlotte Edmond, *What is Open-Source AI and How Could DeepSeek Change the Industry?*, World Econ. F. (February 5, 2025), https://www.weforum.org/stories/2025/02/open-source-ai-innovation-deepseek/.

15 H.R.5586- 118th Congress: DEEPFAKES Accountability Act, (2023).

16 S.2770- 118th Congress: Protect Elections from Deceptive AI Act, (2024).

17 Shelley Shan, *Legislature passes stiffer jail, fine for deepfake fraud,* Taipei Times, (May 17, 2023), https://www.taipeitimes.com/News/front/archives/2023/05/17/2003799936.

18 *Key Issue 5: Transparency Obligations, EU AI Act* https://www.euaiact.com/key-issue/5.

19 Ministry of Electronics and Information Technology (MeitY), Advisory to all Intermediaries to comply with existing IT Rules, (Issued on December 26, 2023).

20 Akshaya Suresh & Neerja Shankar, Revised MeitY Advisory on Deployment of AI Models, (April 22, 2024), https://www.jsalaw.com/newsletters-and-updates/revised-meity-advisory-on-deployment-of-ai-models/.

21 *New advisory of MeitY : AI platforms don't need government permission, focus is on deepfakes,* (March 16, 2024), https://www.businesstoday.in/tech-today/news/story/new-advisory-of-meity-ai-platforms-dont-need-government-permission-focus-on-deepfakes-421703-2024-03-16.

22 Mike Kujawski, *How Adopting Content Provenance Standards Can Help Government Organizations in the Fight Against Mis- and Disinformation,* (November 13, 2024), https://cepsm.ca/how-adopting-content-provenance-standards-can-help-government-organizations-in-the-fight-against-mis-and-disinformation/.

23 *Overview, Coalition for Content Provenance and Authenticity* (C2PA), https://c2pa.org/.

24 Maryam Roshanaie, Mahir R. Khan, *Navigating AI Cybersecurity: Evolving Landscape and Challenges,* JILSA, Vol. 16 (2024).

25 Chris Sledjeski, *Principles for Reducing AI Cyber Risk in Critical Infrastructure: A Prioritization Approach,* MITRE (2023).

26 Cyber Security Agency of Singapore, *Guidelines on Securing AI Systems,* (Issued in October 2024).

27 CyberSecurity and Infrastructure Security Agency, *Shifting the Balance of CyberSecurity Risk: Principles and Approaches for Secure by Design Software* (Issued in October 2023).

28 The issue of whether storing or copying copyrighted material for training generative AI models constitutes infringement has not been addressed in the Report.

29 Sneha Jain and Akshat Agarwal, *AI Training: Transformative Use or Copyright Infringement?,* (July 2, 2024), https://www.medianama.com/2024/07/223-ai-copyright-legal-perspectives-transformative-extractive-uses-copyrighted-works/.

30 Julio Carvalho, *The stubborn memory of generative AI: overfitting, fair use, and the AI Act*, (April 8, 2024), https://copyrightblog.kluweriplaw.com/2024/04/08/the-stubborn-memory-of-generative-ai-overfitting-fair-use-and-the-ai-act/.

31 The Hindu Bureau, *Delhi HC issues summons to OpenAI on ANI's copyright violation plea against ChatGPT,* The Hindu, (November 19, 2024), https://www.thehindu.com/news/national/delhi-hc-issues-summons-to-openai-on-anis-copyright-violation-plea-against-chatgpt/article68885741.ece.

32 *Disclosure:* The Indian Governance and Policy Project (IGAP) is an intervenor in the matter.

33 Clara Chong, *How LLMs Work: Pre-Training to Post-Training, Neural Networks, Hallucinations, and Inference, Towards Data Science,* (February 18, 2025), https://towardsdatascience.com/how-llms-work-pre-training-to-post-training-neural-networks-hallucinations-and-inference/.

34 Pratibha M. Singh, *Evolution of Copyright Law: The Indian Journey,* IJLT, Vol. 16, Issue 2 (2020).

35 Syndicate of the Press of the University of Cambridge v. B.D. Bhandari, 185 (2011) DLT 346.

36 *Copyright Ownership of Generative AI Outputs Varies Around the World, Cooley,* (January 19, 2024), https://www.cooley.com/news/insight/2024/2024-01-29-copyright-ownership-of-generative-ai-outputs-varies-around-the-world.

37 Rich Stim, *Measuring Fair Use: The Four Factors, Stanford Copyright and Fair Use Center,* (April 4, 2013), https://fairuse.stanford.edu/overview/fair-use/four-factors/

38 Authors Guild v. Google Inc, 804 F.3d 202 (2nd Circ., 2015).

39 Andres Guadamuz, *A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs,* GRUR International, Vol. 73, Issue 2, (2024).

40 Super Cassettes v. Hamar Television, 2011 (45) PTC 70 (Del).

41 University of Oxford v. Rameshwari Photocopy Services, (2016) SCCOnline Del 6229.

42 Harsh Kumar, *Employer's Copyright vis-à-vis Author's Right: An Unresolved Legal Dilemma,* JIPR, Vol. 10, (2005).

43 *Supra* note 35.

44 M. Oren Opstein, *et. al, Copyright Office Publishes Report on Copyrightability of AI-Generated Materials,* (February 4, 2025), https://www.skadden.com/insights/publications/2025/02/copyright-office-publishes-report.

45 *EU Policy Questionnaire on the Relationship Between Generative Artificial Intelligence and Copyright and Related Rights,* (January 24, 2025), https://www.jonesday.com/en/insights/2025/01/eu-issues-report-on-relationship-between-generative-ai-and-copyright.

46 Government of United Kingdom, *Copyright and AI: Consultation,* (December 2024), https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence/copyright-and-artificial-intelligence#:~:text=Computer%20generated%20works&text=There%20are%20additional%20issues%20regarding,within%20the%20broader%20copyright%20framework.

47 Nicol Turner Lee, Paul Resnick and Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms,* OECD Publications, (May 22, 2019), https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/12/case-studies-on-the-regulatory-challenges-raised-by-innovation-and-the-regulatory-responses_82fcd441/8fa190b5-en.pdf.

48 Troopers Sanders, *Three Levels of Public Interest AI,* (November 12, 2024), https://www.techpolicy.press/the-three-levels-of-public-interest-ai/.

49 OECD/KDI, *Report on Case Studies on the Regulatory Challenges Raised By Innovation and the Regulatory Responses,* OECD Publishing (2021).

50 Government of United Kingdom, *International Scientific Report on the Safety of Advanced AI* (January 2025).

51 Laura Ellis, *Project Origin Securing Trust in Media,* BBC, (March 4, 2020), https://www.bbc.com/beyondfakenews/trusted-news-initiative/project-origin-securing-trust-in-media.

52 *Supra* note 22.

53 Sharveya Parasnis, *Bank of China Announces Investments Worth 1 Trillion Yuan to Develop AI Industry,* Medianama, (January 25, 2025), https://www.medianama.com/2025/01/223-bank-of-china-announces-1-trillion-yuan-ai-industry-investment/.

54 Bjin Jose, *What is Stargate, Trump's $500-billion bid for global leadership in AI,* The Indian Express, (January 22, 2025), https://indianexpress.com/article/explained/explained-sci-tech/what-is-stargate-trumps-500-billion-ai-project-9793165/.

55 *Cabinet Approves Over Rs 10,300 Crore for IndiaAI Mission, will Empower AI Startups and Expand Compute Infrastructure Access,* PIB Press Release, (March 7, 2024), https://pib.gov.in/PressReleasePage.aspx?PRID=2012375.

56 *Supra* note 9.

57 Leslie D' Monte, *India's Budget will seek to propel $3.7 tn Economy with AI and Tech Investments,* Mint, (January 30, 2025), https://www.livemint.com/budget/expectations/budget-artificial-intelligence-ai-tech-investments-semiconductors-quantum-computing-5g-blockchain-indiaai-mission-11737793559396.html#:~:text=The%20government's%20IndiaAI%20Mission%20has,in%20the%20next%20four%20years.

INDIAN GOVERNANCE AND POLICY PROJECT

The Indian Governance And Policy Project (IGAP) is an emerging think tank focused on driving growth, innovation, and development in India's digital landscape. Specializing in areas like AI, Data Protection, FinTech, and Sustainability, IGAP promotes evidence-based policymaking through interdisciplinary research. By working closely with industry bodies in the digital sector, IGAP provides valuable insights and supports informed decision-making. Core work streams include policy monitoring, knowledge dissemination, capacity development, dialogue and collaboration.

For more details visit: www.igap.in

relations@igap.in | igap.in