To
The Secretary,
Ministry of Electronics and Information Technology,
Electronics Niketan, CGO Complex

Sub: Submission of comments on the Draft Amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021

Dear Sir/Ma'am

The AI Knowledge Consortium (AIKC) welcomes the opportunity to provide comments on Draft Amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ("Draft IT Rules").

Our consortium brings together academic and policy institutions that promote multi-stakeholder participation in AI governance and advocate for inclusive, transparent, and proportionate regulation.

We acknowledge the Ministry of Electronics and Information Technology (MeitY)'s leadership in addressing challenges posed by synthetic and AI-generated content.

Effective enforcement remains one of the greatest challenges in implementing regulations on AI-generated content. The pace, scale, and diversity of digital production make exhaustive compliance monitoring impractical for both government and industry.

In such a context, voluntary disclosure norms and participatory audit mechanisms, anchored in transparency, good-faith reporting, and public accountability, offer a sustainable path forward. We provide related submissions in the annexure. These submissions do not reflect the individual positions of the members of AIKC and have been made as a collective; they are reflective of the opinions of some of our members.

We look forward to remaining engaged in this process.

On behalf of AIKC Members

Secretariat

AI Knowledge Consortium

Annexure - Detailed Submissions

A. Nature of Disclosures

We submit that flexibility to intermediaries to choose their preferred method of disclosing the provenance of synthetically generated information, instead of being bound by the requirements such as that of a specific kind of metadata or mandates to cover a certain percentage of visual display area, may be viable.

The proposed Rule 3(3) requires 'Due Diligence in Relation to Synthetically Generated Information', and mandates the following disclosure obligations for intermediaries offering computer resources enabling creation or modification of synthetically generated information:

- Must ensure such information is labelled or embedded with a permanent unique metadata or identifier;

- *Such label or identifier must be visibly displayed or made audible in a prominent manner on or within the synthetic content, covering at least 10% of the surface area of a visual display or, in the case of audio content, during the initial 10% of its duration; and*

- The label or identifier must enable immediate identification of the content as synthetically generated information. The rule further prohibits intermediaries from modifying, suppressing, or removing such labels or identifiers.

The proposed disclosure obligations does not take into account different types of intermediaries and differing content. The present mandate may present intermediaries with technical challenges, given varying content formats, which may be unsuited to support such labels. Further, experts note that labelling content at scale is unreliable and can be easily removed, particularly in adversarial settings where actors seek to evade detection.[1] Whilst so, labelling *per se* is a welcome move but may have to be implementable without disrupting intermediary services or dissemination of content by content developers / providers.

A prescriptive labelling mandate of synthetic content may not singularly solve the larger issue of misinformation and harmful deepfakes[2]. As more synthetic content emerges in the information space, the complementary safeguard of media literacy for users is also imperative. Studies show that rules mandating AI labels may enhance transparency, but do not significantly change the persuasiveness of the content itself.[3] Therefore, some of our members believe that rather than relying on a mandatory and prescriptive labelling regime, interoperable and verifiable industry-championed disclosure standards may offer a more sustainable approach, with the complementary

[1] Centre for Data Innovation, Why AI-Generated Content Labeling Mandates Fall Short https://www2.datainnovation.org/2024-ai-watermarking.pdf

[2] https://www.hindustantimes.com/opinion/ai-content-regulation-in-india-when-labelling-norms-are-ineffective-101761752385126.html

[3] Stanford University Human-Centered Artificial Intelligence, Labeling AI-Generated Content May Not Change Its Persuasiveness https://hai.stanford.edu/policy/labeling-ai-generated-content-may-not-change-its-persuasiveness

safeguard of media literacy, whilst others advocate distinction in labelling based on content. To cap all labelling under one head would disrupt and harm creative industry for instance.

Disclosure obligations should be consonant with the government's policy objectives in AI. The India AI Governance Guidelines 2025 ("Guidelines") emphasise accountability and understandability.[4] In other words, consumers of online information and regulators should be made aware when the information they consume is synthetically generated, and the method to do so, in a clearly discernible manner may be left to the intermediary or a more viable alternative provided for easy implementation. Entities ought to have the flexibility to implement measures that achieve such objectives. Some studies indicate that prescriptive disclosures risk compliance fatigue and box-ticking.[5]

In India, the Business Responsibility and Sustainability Reporting (BRSR) framework introduced by SEBI is an example of regulatory design that AI-related obligations could adopt. The BRSR shifted corporate disclosures from a fixed questionnaire to principles-based reporting[6], requiring companies to demonstrate how their actions align with policy objectives.

Internationally, frameworks like the EU Corporate Sustainability Reporting Directive (CSRD)[7] and the independent Global Reporting Initiative (GRI) standards[8] similarly emphasise materiality and outcome-based transparency over procedural conformity. These systems make clear what must be achieved (that is, comprehensive, verifiable, and comparable information) without prescribing the exact methodologies, allowing flexibility across sectors and scales of operation.

Flexible yet structured disclosure frameworks have been reported to work well[9]; AI and synthetic-content regulation can benefit from the same philosophy. The Draft IT Rules Amendments should state the purpose of disclosure requirements (enabling users to make informed decisions, allowing regulators to trace provenance, or supporting independent verification) while providing a margin of appreciation for developers and intermediaries to innovate and meet these aims.

For example, one entity may use watermarking, another metadata credentials, and a third public model cards or dashboards, all achieving the same transparency goal. This approach not only

---

[4] Part 1(Key Principles), India AI Governance Guidelines, 2025: available at https://static.pib.gov.in/WriteReadData/specificdocs/documents/2025/nov/doc2025115685601.pdf.

[5] Research and commentary show rules-heavy (form-based) regimes encourage checkbox behavior, whereas principles/outcomes orientation can improve informativeness. For example: Sundvik, Dennis, The Impact of Principles-Based versus Rules-Based Accounting Standards on Reporting Quality and Earnings Management (February 11, 2019). Journal of Applied Accounting Research, Vol. 20, No. 1, 78-93, 2019, Available at SSRN: https://ssrn.com/abstract=3433700.

[6] EY, "BRSR Reporting and the Evolving ESG Landscape in India" (2023), https://www.ey.com/en_in/insights/climate-change-sustainability-services/brsr-reporting-and-the-evolving-esg-landscape-in-india.

[7] "Commission presents voluntary sustainability reporting standard to ease burden on SMEs", https://finance.ec.europa.eu/publications/commission-presents-voluntary-sustainability-reporting-standard-ease-burden-smes_en.

[8] See GRI's Sustainability Reporting Guidelines at https://saiplatform.org/uploads/Modules/Library/GRI-sustainability-reporting-guidelines.pdf.

[9] For example: a 2024 report by the CFA Institute found that under the BRSR framework, "the number of reported data sets has increased annually, the quality of reporting has improved, and there is enhanced attention to detail" (https://rpc.cfainstitute.org/sites/default/files/-/media/documents/article/industry-research/current-state-of-brsr-at-corporate-india.pdf).

promotes compliance diversity and technological innovation but also ensures that regulation evolves alongside rapid changes in generative AI systems.

Norms should be designed to promote compliance capacity across the ecosystem, especially among India's emerging AI startups and MSMEs, who may face high compliance costs with prescriptive mandates. The focus should be on principles-based guidance and capacity-building support, rather than uniform technical obligations.

B. Provenance & labelling methods should use any open, interoperable standard

We suggest that provenance and labelling methods should focus on open, interoperable standards, such as C2PA Content Credentials, IPTC metadata, or W3C verifiable credentials.

The proposed Rule 4(1A) on 'Enhanced Obligations for Significant Social Media Intermediaries(SSMIs)' requires SSMIs to:

- Obtain a user declaration on whether uploaded information is synthetically generated;
- *Deploy reasonable and proportionate technical measures to verify such declarations;*
- Ensure that synthetically generated information is clearly labelled or accompanied by a notice indicating the same

We suggest that provenance and labelling methods may focus on open, interoperable standards, such as C2PA Content Credentials, IPTC metadata, or W3C verifiable credentials. Such standards have been designed to authenticate and preserve the provenance of digital content, and have been found to be technically robust.[10]

Open standards have gained momentum in emerging technology domains (for example, the Alliance for OpenUSD standard has seen industry-wide adoption in the graphics space, or the Physical Security Interoperability Alliance standard with IP-enabled security devices). In general, open standards have been found to improve interoperability, benefit the consumer market[11] and foster innovation.[12]

We therefore submit that reliance may be placed on pre-existing open standards to encourage innovation and harmonization with international provenance standards while avoiding over-reliance on any one provider. MeitY can publish a living registry of accepted standards from which developers can choose what is best suited to their needs, similar to what it has done for cloud service providers.[13]

C. Intermediaries should be incentivised to reduce harms.

[10] For example, the US' NIST Framework for Trustworthy and Responsible AI (2024) (available at https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf) lists the C2PA framework as an example of metadata provenance checking; even the European Commission's blockchain initiative runs off W3C credentials (https://ec.europa.eu/digital-building-blocks/sites/spaces/EBSI/pages/600343491/EBSI%2BVerifiable%2BCredentials).

[11] Zubrinich et al, "Proprietary versus Open Standards", https://www.4ipcouncil.com/application/files/3615/4357/3178/4iP_Council_-_Proprietary-vs-Open-Standards_-_Nov18.pdf.

[12] Macek et al, "Open standard is the new open source", https://pmc.ncbi.nlm.nih.gov/articles/PMC9254132/.

[13] See: Checklist for empanelment readiness, available at https://www.ambud.meity.gov.in.

We recommend that Draft IT Rules explicitly allow intermediaries to pilot and scale proactive harm-reduction measures within clear transparency and accountability parameters.

Kolbel & Kunz (2020) believe that intermediaries possess deep contextual understanding of how their systems operate and consumers behave, including aspects such as model behaviour, user interactions, and evolving threat patterns.[14] This situational knowledge, as enumerated above, possibly places intermediaries in a unique position to design, test, and deploy early-warning mechanisms to indicate if synthetically generated information is leading to user harms. Software and cybersecurity firms have developed a number of measures to mitigate harm, like synthetic-content detectors, provenance tags, and adaptive moderation tools; intermediaries may use them as they deem appropriate.

The nature and intensity of risk varies significantly across intermediaries: large and open networks may amplify the reach and virality of synthetic or misleading content, whereas closed-community platforms, where users have access to dedicated servers or message boards, pose a different moderation and traceability challenge.

Users can create private servers and channels that are largely invisible to outsiders, making it difficult for moderators or regulators to detect harmful or misleading content. The platform's encrypted, community-based structure also limits traceability and rapid intervention once such content spreads within closed groups.

We recommend that the Draft IT Rules should explicitly allow intermediaries to pilot and scale proactive harm-reduction measures within clear transparency and accountability parameters and that the Rules give specific guidance to facilitate the same.

Several global regimes have moved in the direction of letting intermediaries choose how they should respond to threats. The EU Digital Services Act (Articles 34–35)[15] encourages "risk-mitigation innovation" by requiring very large online platforms to assess and mitigate systemic risks using "reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified", subject to external audits. Similarly, the UK's Online Safety Act (2023) allows Ofcom to accept "alternative compliance measures"[16] if platforms can demonstrate equivalent or superior safety outcomes.

India's approach could demonstrate a similar ethos by creating space for intermediary-led experimentation in harm-reduction, including model watermarking pilots, provenance-tool integration (e.g., C2PA, SynthID), algorithmic-impact audits, and user-education campaigns, whilst providing for risk mitigation.

[14] "By collecting, analyzing and combining the data, platform operators can build up an enormous amount of knowledge about the interaction partners, which they are only able to do in their role as intermediaries", as stated in Kolbel & Kunz, "Mechanisms of Intermediary Platforms", available at https://assets.bosch.com/media/global/research/eot/bosch-eot-mechanisms-of-intermediary-platforms.pdf.
[15] Text available at: https://www.eu-digital-services-act.com/Digital_Services_Act_Article_34.html.
[16] Text available at: https://www.legislation.gov.uk/ukpga/2023/50/.