



Government Strengthens Framework to Counter AI-Generated Deepfakes through Legal Safeguards, Platform Accountability and Citizen Protection

13 Responsible AI Projects Approved for Deepfake Detection as Over 11.37 Lakh People Benefit from 6,650 Cyber Security Awareness Workshops

Government Enhances User Safety with Faster Removal of Unlawful AI-Generated Content and Strengthened Grievance Redressal Framework

प्रविष्टि तिथि: 30 JUL 2026 3:15PM by PIB Delhi

The Government remains conscious of the threats posed by deepfakes powered by Artificial Intelligence (AI), including synthetic audio, video and text. With an aim to ensure an open, safe, trusted and accountable cyberspace for users, Government of India has enacted the following laws and rules that address various aspects of the deepfake challenge:

The Information Technology Act, 2000 (“IT Act”)

- Covers offences like identity theft (section 66C), impersonation (section 66D), privacy violations (section 66E), publishing or transmitting obscene or sexually explicit content (sections 67, 67A)
- Provision to issue blocking orders to intermediaries for blocking access to specific information/ link (section 69A)
- Provision to issue notice to intermediaries for removal of information being used to commit unlawful act (section 79)
- Besides, the Act also empowers Police to investigate offences (Section 78 and 80).

The Bharatiya Nyaya Sanhita, 2023 (“BNS”)

- Section 353 aims to curb the spread of misinformation and disinformation by penalizing the act of making false or misleading statements, rumours, or reports that can cause public mischief or fear.
- Organised cybercrimes involving deepfake content can also be prosecuted under section 111.

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (“IT Rules”)

The IT Rules cast specific obligations on intermediaries to observe due diligence while discharging their duties and shall inform the users of computer resources not to host, display, upload, modify, publish, transmit, update or share any information that is

- obscene, pornographic, invasive of privacy, or promotes hate or violence;
- harms children;
- misleads or deceives, including through deep-fakes;
- impersonates others, including via AI;
- threatens national security or public order;
- violates any applicable law.

Intermediaries must clearly inform users through terms of service and user agreements about the consequences of sharing unlawful content, including content removal, account suspension, or termination.

Social media intermediaries having 50 lakhs or above registered user base in India (SSMIs) are required to comply with following additional obligations:

- SSMIs offering messaging services must help law enforcement trace originators of serious or sensitive content.
- To use automated tools to detect and limit spread of unlawful content.
- Publish compliance reports, appoint local officers, and share physical address based in India for compliances and law enforcement coordination.
- Offer voluntary user verification, internal appeals, and fair hearing before taking suo-moto action.

Grievance Redressal Mechanism:

Intermediaries are required to appoint Grievance Officers and resolve complaints within the prescribed timelines. Users can appeal online at <https://www.gac.gov.in> if their complaints are not addressed by the intermediaries' Grievance Officers. GACs ensure accountability and transparency of content moderation decisions.

The complaints registered with the Police are investigated by the concerned Law Enforcement Agencies (LEAs) under the relevant statutory framework. Since 'Police' and 'Public Order' are State subjects as per the Seventh Schedule of the Constitution of India, the States/UTs are primarily responsible for the prevention, detection, investigation and prosecution of crimes including cybercrime through their Law Enforcement Agencies (LEAs).

Recent Amendments:

On 10th February, 2026, the Government strengthened the regulatory framework by amending the IT Rules, 2021 to address harms arising from synthetically generated information (SGI), including deepfakes and AI-generated content.

Key points related to the amendment are as follows: -

- Intermediaries are required to ensure clear labelling and traceable metadata for permissible AI-generated content, so that users can easily identify synthetically generated material and prevent deception or misuse.
- It further strengthens user accountability and platform due diligence, including mandatory user awareness regarding legal consequences of unlawful AI-generated content and stronger compliance obligations for social media intermediaries.
- Importantly, the guidelines explicitly cover child sexual exploitation material, non-consensual intimate imagery, impersonation and other harmful AI-generated content, requiring platforms to prevent such content and take prompt action when detected.
- Strengthening of timelines for compliance, including reduced timelines for removal of unlawful information upon actual knowledge upon valid reasoned intimation from the Appropriate Government or court orders (timeline reduced from 36 hours to 3 hours) and for grievance redressal

(including special categories such as nudity/impersonation etc.) (timeline reduced from 72 hours to 36 hours and 24 hours to 2 hours for sensitive matters, respectively).

- Intermediaries are obligated to deploy reasonable and appropriate technical measures, including automated tools or other suitable mechanisms, to not allow any user to create, generate, modify, alter, publish, transmit, share, or disseminate, as the case may be, any synthetically generated information that violates any law for the time being in force.
- IT Rules mandates the Significant Social Media Intermediaries (SSMIs) to take reasonable efforts to deploy appropriate technical measures, including automated tools or other suitable mechanisms, to proactively identify information that depicts any act or simulation in any form depicting rape, child sexual abuse or conduct, whether explicit or implicit, or any information which is exactly identical in content to information that has previously been removed.

In case of failure of the intermediaries to observe the legal obligations as provided in the IT Rules, they lose their exemption from third party information provided under section 79 of the IT Act. They are liable for consequential action or prosecution as provided under any extant law.

Advisories and Standard Operating Procedure (SOP):

The government has issued multiple advisories to the intermediaries, including the Social Media Intermediaries, emphasizing the observance of due diligence obligations under the IT Act and IT Rules. The details of the advisories issued and SOPs formulated are as follows:

- Vide Advisory dated 29.12.2025, government reiterated to the intermediaries regarding observance of statutory due diligence obligations under the IT Act and IT Rules, for preventing hosting, publication, transmission, sharing or uploading of vulgar, indecent, obscene, pornographic and other unlawful content on their platforms. It also advised the intermediaries to undertake an immediate review of their internal compliance frameworks, content moderation practices and user enforcement mechanisms, and to ensure strict and continuous adherence to the provisions of the IT Act and the IT Rules.
- An advisory dated 09.02.2026 was issued to intermediaries regarding the responsible handling of information relating to religious matters. The intermediaries were advised to urgently deploy enhanced monitoring and accountability frameworks to ensure that religious content is handled with due sensitivity, accuracy and respect, and that no distorted, misleading or unlawful religious material is permitted to proliferate on their platforms. And that any non-compliance by the intermediaries with the statutory due diligence obligations may result in loss of exemption provided under section 79 of the IT Act and exposure to consequential legal action under applicable laws.
- An advisory dated 16.03.2026 was issued to intermediaries with respect to the generation, hosting, publication, transmission, sharing or uploading of abusive, defamatory, objectionable, derogatory and misleading synthetically generated information.
- A Standard Operating Procedure (“SoP”) to curtail dissemination of Non-Consensual Intimate Imagery (NCII) content on online platforms has been formulated and released on 11.11.2025. The SoP provides detailed guidance for victims, intermediaries and law enforcement agencies to ensure prompt and uniform action against the online dissemination of NCII content including intimate or morphed images shared without consent.

Indigenous AI-based tools for preserving India’s cultural heritage

Gyan Bharatam Web Portal uses AI (Artificial Intelligence) -enabled technologies, including Large Language Script Models (LLSM), Optical Character Recognition (OCR), Handwritten Text Recognition (HTR), script recognition, semantic search and machine-assisted translation, to provide intelligent access to India’s manuscript heritage.

Gyan Mitram, the AI-powered knowledge interface of the Gyan Bharatam Web Portal uses LLSM, OCR, HTR, script recognition, semantic search and machine-assisted translation to enable intelligent interpretation, contextualization and exploration of manuscripts by integrating manuscript images, transcriptions and translations.

BharatiyaGPT, features “Manuscript-to-Conversational AI” on Gyan Bharatam Web Portal converting manuscript heritage into AI-ready, machine-readable knowledge and provides a multilingual and multimodal access, including text, audio and voice-based interaction through its “Talk to Manuscript” feature, thereby making India’s knowledge traditions more accessible to scholars, students, researchers and the public.

Detection of Deepfake Content

The Safe & Trusted AI pillar of the IndiaAI Mission aims to promote the responsible development, deployment and adoption of AI through indigenous governance frameworks, standards, tools and evaluation mechanisms.¹³ Responsible AI projects have been approved in educational institutions across the country including deepfake detection.

Key initiatives related to deepfake detection include:

- Saakshya, multi-agent framework developed by IIT Jodhpur & IIT Madras for deep fake detection
- AI Vishleshak for improving Audio-Visual forgery detection system
- IIT Kharagpur’s project on Real-Time Voice Deepfake Detection System _

Awareness campaigns:

Information Security Education and Awareness (ISEA)- Government is implementing a project on ‘Information Security Education and Awareness (ISEA)’ for generating human resources in Information Security and creating general awareness on various aspects of cyber hygiene & cyber security among the masses. So far, 6,650 awareness workshops have been conducted across the country covering 11.37 lakh+ participants, including school/colleges students, teachers, law enforcement, government personnel, and general public.

Further, multilingual awareness material in the form of handbooks, short videos, posters, brochures, cartoon stories for children, etc. published and disseminated through print, electronic, social media and www.isea.gov.in & <https://staysafeonline.in/>.

CERT-In: The Indian Computer Emergency Response Team also regularly issues guidelines on AI-related threats and countermeasures, including deepfake ;shares safety and security tips and awareness posters, info-graphics and videos on its official websites and social media handles and is aimed at sensitizing internet users on cyber security attacks and frauds including online safety measures for children. It has published an advisory in November 2024 on deepfake threats and measures that need to be followed to stay protected against deepfakes.

This information was submitted by Union Minister of State Shri Jitin Prasada in Lok Sabha on 29.07.2026.

Vinod Kumar / Kanishk Sharma

(रिलीज़ आईडी: 2291671) आगंतुक पटल : 2300

इस विज्ञप्ति को इन भाषाओं में पढ़ें: हिन्दी , Gujarati , Telugu , Kannada

