
FROM TAKEDOWN TO TRIAL: PRESERVING SYNTHETIC-MEDIA EVIDENCE IN INDIA

Dr. Abhishek Kumar Verma, Assistant Professor, School of Law, TS Mishra University,
Lucknow

Mr. Vallabh Srivastava, LLM, School of Law, TS Mishra University, Lucknow

ABSTRACT

India's 2026 amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules create a new legal architecture for synthetically generated information (SGI). The regime combines rapid removal duties, user declarations, technical verification, labels and permanent provenance metadata with an existing obligation to preserve removed information and associated records for 180 days without vitiating the evidence. At the same time, the Bharatiya Sakshya Adhiniyam, 2023 (BSA) makes electronic records admissible through section 63 and a certificate framework that records device particulars and hash values. This article identifies a largely overlooked problem at the intersection of these regimes: what happens to synthetic-media evidence after a platform disables public access but before a court evaluates authenticity? It argues that India faces a takedown-preservation paradox. Rapid removal protects victims and the information ecosystem, but a file-only preservation practice may discard the context, provenance, declarations, verification results and platform logs needed to prove or contest authenticity. A hash can establish post-capture integrity; it cannot establish that the depicted event was real or that the file was not already synthetic when hashed. Drawing on Indian law, recent Supreme Court doctrine, the EU AI Act, Federal Rule of Evidence 901, NIST guidance and the C2PA provenance model, the article proposes a Preserve-Before-Takedown (PBT) protocol. The protocol separates public removal from evidentiary destruction and requires a standardized evidence bundle containing native media, contextual capture, provenance metadata, moderation events, user declarations, verification outputs, cryptographic hashes and chain-of-custody information. It also proposes targeted rule-making and judicial practice directions so that synthetic-media authenticity remains transparent, testable and contestable by both prosecution and defence. The article's central claim is institutional: intermediary compliance and evidentiary reliability should not operate as separate legal silos. Preservation duties should be designed at the moment of moderation, so courts receive verifiable records rather than stripped media files. The proposed framework therefore links platform governance, electronic-

evidence certification, forensic examination and fair-trial rights in a single preservation workflow.

Keywords: synthetically generated information; deepfakes; electronic evidence; Bharatiya Sakshya Adhiniyam; intermediary liability; provenance; chain of custody; artificial intelligence.

I. INTRODUCTION

Synthetic media has changed the evidentiary meaning of seeing and hearing. A video can now be photorealistic, internally coherent and emotionally persuasive even when the depicted act never occurred. Deepfake scholarship initially focused on privacy, reputational harm, democratic discourse and national security; more recent work has turned to the courtroom problem of authenticating media that can be fabricated at scale.¹

India's position materially changed in February 2026. Through G.S.R. 120(E), the Central Government amended the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules) to define 'synthetically generated information' and impose specific due-diligence duties concerning its creation, dissemination, labelling and provenance.² The amendments also reduce the principal actual-knowledge compliance window in Rule 3(1)(d) from thirty-six hours to three hours, while certain impersonation-related grievances must be acted upon within two hours.³

That speed is understandable. Harmful synthetic media can replicate before formal adjudication catches up. Yet speed produces a second-order legal problem: the same content that must disappear from public view may later become central evidence in criminal prosecution, civil litigation, regulatory enforcement or a defence case. The IT Rules already recognize this danger. Rule 3(1)(g) requires an intermediary that removes information or disables access to preserve the information and associated records for 180 days, or longer when lawfully required, 'without vitiating the evidence'.⁴ Section 67C of the Information Technology Act, 2000

¹ Bobby Chesney and Danielle Keats Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security', 107 California Law Review 1753 (2019); Rebecca Delfino, 'Deepfakes on Trial: A Call to Expand the Trial Judge's Gatekeeping Role to Protect Legal Proceedings from Technological Fakery', 74 Hastings Law Journal 293 (2023).

² Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E), Gazette of India, Extraordinary, Pt. II, sec. 3(i), Feb. 10, 2026 (in force from Feb. 20, 2026).

³ Id., rule 3(a)(ii)(A) and rule 3(b)(ii), amending Rule 3(1)(d) and Rule 3(2)(b) of the IT Rules, 2021.

⁴ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, rule 3(1)(g), as amended.

separately authorizes prescription of preservation and retention duties for intermediaries.⁵

The evidentiary destination of such preserved material is the Bharatiya Sakshya Adhiniyam, 2023 (BSA), which came into force on 1 July 2024.⁶ Sections 61 to 63 recognize electronic and digital records and prescribe the conditions under which computer output may be admitted. The Schedule to the BSA also gives a structured certificate format that records device information and hash values. These are important integrity safeguards. They are not, however, a complete answer to synthetic-media authenticity.

This article develops a distinction that should become central to Indian electronic-evidence doctrine: integrity is not authenticity, and authenticity is not truth. A cryptographic hash can show that a particular file remained unchanged after a specified point in time. A section 63 certificate can establish the conditions and manner in which a computer output was produced. Neither fact necessarily proves that a person shown in a video performed the depicted act, that a voice belongs to the apparent speaker, or that the media was not already AI-generated before it entered the chain of custody. In synthetic-media disputes, courts therefore need evidence about provenance before and around capture, not merely integrity after capture.

The article uses a doctrinal and comparative method. It reads the BSA together with the Information Technology Act and the IT Rules as an integrated evidentiary-regulatory system; examines Supreme Court doctrine on electronic evidence; and compares the Indian framework with the European Union's machine-readable marking requirement, the authentication logic of Federal Rule of Evidence 901, and technical provenance approaches represented by NIST and the Coalition for Content Provenance and Authenticity (C2PA).⁷

II. THE 2026 REGULATORY SHIFT: FROM CONTENT MODERATION TO PROVENANCE GOVERNANCE

A. A statutory category for synthetic media

The 2026 amendment defines SGI as audio, visual or audio-visual information artificially or

⁵ Information Technology Act, 2000, s. 67C.

⁶ Bharatiya Sakshya Adhiniyam, 2023, Act No. 47 of 2023, s. 1; Notification S.O. 849(E), Feb. 23, 2024.

⁷ Regulation (EU) 2024/1689 (Artificial Intelligence Act), art. 50; Federal Rules of Evidence, rule 901; Bilva Chandra et al., Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency, NIST AI 100-4 (2024); C2PA, Content Credentials: Technical Specification, version 2.4 (2026).

algorithmically created, generated, modified or altered through a computer resource so that it appears real, authentic or true and depicts an individual or event in a manner likely to be perceived as indistinguishable from a natural person or real-world event.⁸ The definition is deliberately narrower than all AI-assisted content. It excludes routine good-faith editing, formatting, enhancement, technical correction, colour adjustment, noise reduction, transcription and compression that do not materially alter meaning; it also excludes ordinary creation of documents, presentations, PDFs, educational or research outputs where no false document or false electronic record results, and accessibility or clarity improvements that do not manipulate a material part of the underlying audio-visual information.⁹

This calibration matters evidentially. Courts should not equate an AI-assisted workflow with deceptive synthetic evidence. The legal inquiry should remain claim-specific: what does the proponent say the item proves, and what transformation occurred before the item was offered? A transcript produced by speech-to-text may raise accuracy questions, but it is not necessarily SGI. A cloned voice purporting to be the accused's confession presents a fundamentally different authenticity problem.

B. Labels, permanent metadata and platform-side verification

Rule 3(3) creates a provenance-oriented duty for intermediaries that offer computer resources capable of creating or facilitating SGI. Unlawful categories must be addressed through reasonable and appropriate technical measures. For permissible SGI, the intermediary must provide a prominent label or audio disclosure and, to the extent technically feasible, embed permanent metadata or another appropriate technical provenance mechanism, including a unique identifier capable of identifying the intermediary's computer resource used to create or alter the information.¹⁰ The intermediary must not enable modification, suppression or removal of the label or the embedded provenance information.¹¹

Significant social media intermediaries receive an additional ex ante role. Before publication they must obtain a user declaration about whether information is synthetically generated, deploy appropriate technical measures to verify the declaration having regard to the nature,

⁸ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, rule 2(a)(ii), inserting Rule 2(1)(wa).

⁹ Id.

¹⁰ Id., rule 3(c), inserting Rule 3(3)(a).

¹¹ Id., Rule 3(3)(b).

format and source of the information, and label the content where the declaration or technical verification confirms its synthetic character.¹² These obligations create evidence that did not previously exist at scale: the user's declaration, the platform's verification result, the time of verification, the technical method or model version used, and the platform's labeling decision. Each may later be relevant to authenticity, knowledge, intent, attribution or notice.

C. Removal is not supposed to mean evidentiary destruction

The same amendments explicitly contemplate account suspension or termination 'without vitiating the evidence' for certain SGI violations.¹³ That phrase should be read alongside Rule 3(1)(g)'s 180-day preservation duty. The architecture is therefore not a simple notice-and-delete regime. It is more accurately a notice-remove-preserve regime.

The problem is that 'preserve' remains underspecified. A platform can technically preserve a copy of the media file yet still lose or fail to export the context necessary for legal authentication: the original upload timestamp, post identifier, edit history, user declaration, moderation queue data, provenance manifest, unique SGI identifier, account linkage, associated captions, comments establishing context, or records of automated verification. A file-only preservation practice may satisfy the intuition that 'the content still exists', while frustrating the evidentiary purpose of the rule.

III. THE BSA AUTHENTICATION GAP

A. What section 63 does well

The BSA treats electronic and digital records as documents and gives them the same legal effect as other documents, subject to the special proof mechanism in section 63.¹⁴ Section 63 permits a computer output to be admitted if statutory conditions concerning regular use, ordinary-course input, proper operation and derivation of information are satisfied. A certificate must identify the electronic record, describe the manner of production, provide relevant device particulars and address the statutory conditions.¹⁵

¹² Id., rule 4(a), inserting Rule 4(1A).

¹³ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, Rule 3(1)(ca)(ii)(II), as inserted by G.S.R. 120(E).

¹⁴ Bharatiya Sakshya Adhiniyam, 2023, ss. 2(1)(d), 61-63.

¹⁵ Id., s. 63(2)-(4).

The Supreme Court's electronic-evidence jurisprudence under section 65B of the Indian Evidence Act, 1872 remains highly relevant because section 63 substantially carries forward the certificate structure. In *Anvar P.V. v. P.K. Basheer*, the Court treated the statutory certificate as the route for proving computer output, and *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal* reaffirmed that the certificate is a condition precedent to admissibility where the provision applies.¹⁶ Most recently, in *Pooranmal v. State of Rajasthan*, the Supreme Court expressly linked the older section 65B requirement with section 63 of the BSA and declined to rely upon call detail records where the mandatory certificate had not been proved.¹⁷

The BSA also recognizes expert assistance. Section 39(2) makes the opinion of an Examiner of Electronic Evidence under section 79A of the Information Technology Act relevant when a court must form an opinion on information transmitted or stored in a computer resource or other electronic or digital form.¹⁸ This is significant for deepfake detection, but expert opinion remains downstream: an expert cannot meaningfully reconstruct platform context that was never preserved.

B. Hashes prove stability, not historical truth

The BSA's certificate form expressly contemplates hash values. That is a welcome modernization because hashing allows the court and parties to test whether the exhibit offered later is bit-for-bit identical to the file that was hashed earlier.¹⁹ But the inference has limits. Suppose a fabricated video is generated at 10:00 a.m., uploaded at 10:05 a.m., downloaded by an investigator at 11:00 a.m., hashed at 11:01 a.m., and never altered thereafter. The hash establishes excellent post-11:01 integrity. It says nothing about whether the scene was genuine at 10:00 a.m.

This difference can be stated in three layers. First, integrity asks whether the bits have changed after a reference point. Second, provenance asks where the item came from, what transformations preceded the reference point, and which systems or actors handled it. Third, semantic authenticity asks whether the item is what the proponent claims it to be - for example,

¹⁶ *Anvar P.V. v. P.K. Basheer*, (2014) 10 SCC 473; *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*, (2020) 7 SCC 1.

¹⁷ *Pooranmal v. State of Rajasthan*, 2026 INSC 217, paras 48-52.

¹⁸ *Bharatiya Sakshya Adhiniyam*, 2023, s. 39(2); *Information Technology Act*, 2000, s. 79A.

¹⁹ *Bharatiya Sakshya Adhiniyam*, 2023, Schedule to s. 63(4), certificate form for electronic records (including hash particulars).

a genuine recording of an event rather than a synthetic reconstruction. These layers overlap, but they are not interchangeable.

Pooranmal itself indirectly supports the distinction. The Court reiterated the proposition, drawn from *Anvar*, that compliance with the electronic-record proof requirement precedes the separate question of genuineness.²⁰ Synthetic media makes that second question more demanding because visual plausibility no longer supplies a dependable shortcut to authenticity.

C. The 'liar's dividend' enters evidence law

Deepfakes create a dual risk. Fabricated evidence may be believed because it looks real; authentic evidence may be falsely dismissed as AI-generated because fabrication is now plausible. Scholars describe the latter dynamic as a 'liar's dividend': the existence of deepfakes gives wrongdoers a new vocabulary for denying genuine recordings.²¹ A procedural framework that demands impossible proof of every digital file may therefore be as dangerous as a framework that accepts visually persuasive media too readily.

The correct response is not a categorical presumption that labelled content is fake, that unlabelled content is genuine, or that a detection model decides authenticity. Labels can be removed outside controlled systems; provenance metadata can be stripped through re-encoding, screenshots or analogue capture; detection models can produce false positives and false negatives. NIST's work on synthetic-content transparency accordingly treats provenance, watermarking, detection and related techniques as complementary rather than interchangeable solutions.²²

IV. THE TAKEDOWN-PRESERVATION PARADOX

A. Three hours may be enough to remove a post, but not to reconstruct an evidentiary record

Rule 3(1)(d)'s three-hour compliance window prioritizes rapid mitigation once the legally specified form of actual knowledge is received.²³ The problem arises if a platform's operational

²⁰ *Pooranmal v. State of Rajasthan*, 2026 INSC 217, paras 49-52, discussing *Anvar P.V.* and *Arjun Panditrao*.

²¹ Bobby Chesney and Danielle Keats Citron, *supra* note 1, at 1785-1786.

²² Bilva Chandra et al., *Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency*, NIST AI 100-4 (2024).

²³ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, amending Rule 3(1)(d) by substituting 'within three hours' for 'thirty-six hours'.

workflow treats content removal as the terminal action. In a synthetic-media case, the evidentiary object is not merely the visible video or audio. It may include the relationship between the media and its provenance record, the identity of the upload event, the user's declaration, the verification output, and the moderation actions generated by the complaint.

A rapid response can therefore protect one legal interest while impairing another. A victim has a strong interest in prompt removal. The prosecution may later need the original platform-side record to prove fabrication or attribution. The accused may need the same material to demonstrate that the exhibit introduced by the prosecution is a derivative copy, that the platform's classifier was uncertain, that metadata was absent at upload, or that the alleged account was compromised. A preservation rule must serve both directions of proof.

B. The disappearing context problem

Platform evidence is relational. A native file detached from the platform may not retain the webpage context in which it appeared. A screenshot can preserve visual context but usually loses hidden metadata and may itself be manipulated. A screen recording can preserve interface behaviour but not necessarily server-side timestamps. An API export can be structured and precise but may omit what a user actually saw. Forensic preservation must therefore be layered rather than format-monolithic.

At a minimum, the legally relevant context for disputed SGI may include: the native media file; upload and modification timestamps; platform post or message identifiers; the account identifier and account-state information; the user SGI declaration; the platform's technical-verification output; the label shown to users; embedded provenance metadata and unique identifiers; hashes; captions and surrounding text; the moderation event log; the complaint or notice that triggered action; and information showing whether the media was transcoded, resized or re-encoded by the platform. The exact bundle should be proportionate to the dispute, but the preservation system should be capable of retaining these categories.

C. Retention without a schema is not enough

Rule 3(1)(g)'s 180-day requirement is valuable, but duration is only one dimension of preservation. Evidentiary usability also depends on scope, format, integrity, documentation, access controls and disclosure. Retaining an opaque database object that cannot later be exported with intelligible field definitions is not equivalent to preserving court-ready evidence.

Nor is retaining only the final transformed copy if the platform has discarded the original upload or provenance manifest.

The phrase 'without vitiating the evidence' should therefore be treated as imposing a functional standard. Preservation should maintain enough of the record's native and contextual attributes to permit a court, an opposing party and an independent expert to examine its integrity and provenance. Where the platform transforms user media as part of ordinary processing, both the received version and the platform-delivered version should be preserved when material differences may matter.

D. Fair trial requires contestable authenticity

The BSA describes itself as legislation for general rules and principles of evidence for fair trial.²⁴ Fair trial in digital cases cannot mean that one party receives only a conclusory platform statement - 'our system detected synthetic content' - while the underlying evidentiary basis is unavailable for challenge. The Supreme Court has repeatedly treated fairness of procedure as integral to Article 21, and has emphasized that a criminal trial is not merely a contest for conviction but a process for discovering truth through fair opportunity.²⁵

This does not require disclosure of proprietary source code in every case. It requires proportional contestability. A court may use confidentiality orders, in-camera inspection, neutral experts or limited technical disclosures. But where a platform's verification result materially influences attribution or authenticity, the affected party should receive enough information to understand what was tested, when it was tested, the version of the system used, the result produced, and known limitations relevant to that result.

V. COMPARATIVE AND TECHNICAL LESSONS

A. European Union: machine-readable marking as infrastructure

Article 50(2) of the EU AI Act requires providers of AI systems that generate synthetic audio, image, video or text to ensure that outputs are marked in a machine-readable format and detectable as artificially generated or manipulated, subject to stated exceptions and technical

²⁴ Bharatiya Sakshya Adhiniyam, 2023, long title.

²⁵ Maneka Gandhi v. Union of India, (1978) 1 SCC 248; Zahira Habibullah Sheikh v. State of Gujarat, (2004) 4 SCC 158.

feasibility.²⁶ Article 50(4) separately imposes disclosure duties for deepfakes in specified circumstances.²⁷ The significance for India is not that EU law should be transplanted wholesale, but that transparency is treated as technical infrastructure rather than merely an on-screen warning.

India's 2026 Rules already move in the same direction by requiring permanent metadata or other provenance mechanisms to the extent technically feasible. The missing evidentiary step is to ensure that those machine-readable signals survive moderation action and can be exported in a verifiable form for legal proceedings.

B. United States: authentication focuses on the proponent's claim

Federal Rule of Evidence 901 asks the proponent to produce evidence sufficient to support a finding that an item is what the proponent claims it is. The rule offers multiple non-exclusive methods, including witness knowledge, distinctive characteristics, voice identification and evidence describing a process or system and showing that it produces an accurate result.²⁸ This claim-relative model is useful for synthetic media. The proof necessary to establish 'this is the exact file downloaded from platform X' differs from the proof necessary to establish 'this video genuinely records person Y committing act Z'.

American scholarship has proposed stronger judicial gatekeeping for deepfake authenticity because conventional authentication may be too permissive when fabrication is technologically sophisticated.²⁹ Indian law need not reproduce the Federal Rules structure, but it can adopt the underlying discipline: courts should identify the precise authenticity proposition and demand evidence tailored to that proposition.

C. C2PA and the value - and limits - of cryptographic provenance

C2PA's Content Credentials model provides a technical framework for storing and validating provenance assertions, digitally signed claims and content bindings. Its specification describes 'authenticity' as a property grounded in facts such as provenance data and hard bindings that

²⁶ Regulation (EU) 2024/1689 (Artificial Intelligence Act), art. 50(2).

²⁷ *Id.*, art. 50(4).

²⁸ Federal Rules of Evidence, rule 901(a)-(b).

²⁹ Rebecca Delfino, *supra* note 1; Rebecca Delfino, 'Deepfakes on Trial 2.0: A Revised Proposal for a New Federal Rule of Evidence to Mitigate Deepfake Deceptions in Court' (2025), SSRN working paper.

can be cryptographically verified as not having been tampered with.³⁰ Hard bindings may use cryptographic hashes to associate a manifest with an asset.³¹

C2PA is not a legal truth machine. A valid credential can establish that a particular signer made a particular assertion and that bound content has not been tampered with in specified ways; it cannot guarantee that every assertion is truthful or that content lacking a credential is false. Its evidentiary value lies in making parts of the provenance chain testable. That is precisely why legal preservation should capture provenance credentials when they exist rather than stripping them during download, transcoding or moderation.

VI. A PRESERVE-BEFORE-TAKEDOWN (PBT) PROTOCOL FOR INDIA

The central reform proposed here is procedural rather than prohibitory. Before an intermediary removes or disables public access to content that is reasonably likely to become evidence of an SGI-related violation, it should create a standardized Preserve-Before-Takedown (PBT) evidence bundle. Public access can then be disabled immediately; preservation occurs in a segregated evidence environment. The protocol should operate automatically where feasible so that evidentiary preservation does not become an excuse for delaying victim protection.

A. Step 1: trigger and scope

The PBT trigger should arise when removal occurs under Rule 3(1)(d), Rule 3(1)(cb), Rule 3(2)(b) or another provision involving alleged impersonation, false electronic records, non-consensual synthetic content or materially disputed SGI. The system should avoid collecting unrelated data. Preservation should be tailored to the content item, relevant account events and the period reasonably necessary to establish creation, upload, transformation, notice and moderation.

B. Step 2: native media and contextual capture

The platform should preserve the native or earliest available uploaded file, the version delivered to users, and a contextual representation of the post or message as displayed immediately before takedown. Where the platform has transformed the file, it should record

³⁰ Coalition for Content Provenance and Authenticity (C2PA), Content Credentials: Technical Specification, version 2.4 (2026), definitions of provenance, authenticity and content binding.

³¹ *Id.*

the transformation or transcoding steps material to later comparison. Each preserved object should receive a stable evidence identifier.

C. Step 3: provenance packet

The provenance packet should contain any SGI label, embedded metadata, unique identifier, C2PA or comparable credential, watermark-detection result and other technical provenance information available to the intermediary. For significant social media intermediaries, it should also preserve the user's SGI declaration and the platform's technical-verification result required by Rule 4(1A).³²

D. Step 4: moderation and event log

A chronology should record the relevant upload, edit, declaration, label, verification, complaint, actual-knowledge notice, automated moderation and takedown events. Logs should be exported in a human-readable form with field definitions and, where technically practicable, in a machine-readable form for independent analysis. This chronology can later assist the court in evaluating chain of custody and whether a derivative copy accurately represents the platform record.

E. Step 5: cryptographic sealing

The native file, delivered file, provenance packet and log export should be individually hashed using a currently secure algorithm and then associated through a signed manifest or equivalent evidence package. The BSA Schedule currently lists common hash options, but operational guidance should prefer contemporary collision-resistant algorithms such as SHA-256 or stronger unless a legacy system makes another algorithm necessary. A legacy hash may be recorded for compatibility, but it should not be the sole integrity control where a stronger hash is available.

F. Step 6: BSA-ready certification

The platform should maintain the information required to support a section 63 certificate if the record is later produced in court. Because section 63(4) now requires a certificate at each

³² Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, inserting Rule 4(1A).

instance where the electronic record is submitted for admission, evidence-export systems should be designed to generate a reproducible record of the device/system particulars, method of production and relevant statutory conditions.³³ The evidentiary export should not falsely certify semantic authenticity; it should certify what the platform can actually know - source system, acquisition method, timestamps, transformations, preserved metadata and integrity checks.

G. Step 7: retention, legal hold and deletion discipline

The baseline retention period should follow Rule 3(1)(g)'s 180 days. A litigation hold should override routine deletion when a court, competent government agency or properly recognized legal process requires longer preservation.³⁴ At the end of the lawful retention period, evidence should be securely deleted unless another legal basis applies. Preservation should not become indefinite surveillance by default.

H. Step 8: controlled disclosure and defence access

When evidence is produced, the receiving authority should obtain the evidence bundle, its manifest and the certification necessary under the BSA. In criminal cases, material reasonably capable of undermining the prosecution's authenticity theory or supporting the defence should not be filtered out merely because it originated in a platform's moderation system. Courts should be able to order production of the relevant verification outputs, provenance records and audit information, subject to proportionate protections for privacy, cybersecurity and trade secrets.

VII. DOCTRINAL AND INSTITUTIONAL REFORMS

A. A preservation standard under the IT Rules

MeitY should issue a narrowly framed standard operating procedure under the IT Rules specifying what 'without vitiating the evidence' requires for SGI. The SOP should define a minimum evidence schema rather than mandating a particular vendor or proprietary technology. The schema should cover: media objects, contextual capture, timestamps,

³³ Bharatiya Sakshya Adhiniyam, 2023, s. 63(4).

³⁴ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, rule 3(1)(g).

identifiers, user declarations, verification outputs, labels, provenance metadata, relevant moderation events, cryptographic hashes and chain-of-custody information.

This is consistent with the existing legal architecture. Rule 3(1)(g) already imposes preservation; the 2026 Rules already create provenance and verification data; and the BSA already asks for structured information about electronic-record production. The reform would connect obligations that presently operate in separate compliance silos.

B. Judicial practice directions for disputed synthetic media

High Courts should consider practice directions for cases in which a party specifically disputes whether audio-visual evidence is synthetic or materially AI-altered. The directions need not create a new exclusionary rule. They can require an early authenticity conference at which the court identifies: (i) what the proponent claims the item is; (ii) whether section 63 applies and the certificate is complete; (iii) what provenance material exists; (iv) whether the original platform-side record has been preserved; (v) whether expert examination is required; and (vi) what controlled disclosures are necessary for a fair challenge.

A court should also distinguish between absence of provenance and evidence of fabrication. The fact that a viral clip lacks a C2PA credential, watermark or platform label should not automatically make it inadmissible or false. Many genuine legacy recordings have no such signals. Conversely, presence of a valid provenance credential should carry weight only to the extent the signer, assertions and bindings are relevant to the proposition being proved.

C. Standards for expert evidence and AI detectors

Expert evidence on deepfake detection should disclose the method used, version, test conditions, known error characteristics where available, whether the model has been evaluated on media comparable to the disputed item, and whether compression or platform transcoding may affect performance. The court should resist binary testimony that an opaque model has 'proved' a file fake. Detection is probabilistic evidence that should be integrated with provenance, contextual facts and chain-of-custody material.

Section 39(2) of the BSA provides a statutory basis for the relevance of an Examiner of

Electronic Evidence, but institutional capacity will matter.³⁵ Training should therefore extend beyond traditional disk and device forensics to synthetic-media provenance, watermarking, generative editing traces and the interpretation of content credentials.

D. Data minimization and privacy safeguards

A preservation system can itself become a rights risk if platforms indiscriminately retain users' communications, identifiers and behavioural data. The PBT protocol should therefore preserve only material reasonably connected to the disputed content and the legal purpose for which preservation is required. Access should be logged, role-limited and auditable; evidence stores should be segregated from ordinary advertising or recommendation systems; and retention should end when the lawful basis expires.

The key is separation of functions: takedown protects the public-facing information environment; evidence preservation protects adjudication; privacy safeguards constrain the preserved corpus. Treating those functions distinctly avoids the false choice between rapid harm mitigation and reliable proof.

VIII. A MODEL LEGAL TEST FOR COURTS

For practical use, Indian courts can approach disputed synthetic-media evidence through five sequential questions. First, relevance: what fact in issue does the media purport to prove? Second, admissibility: have the applicable requirements of section 63 of the BSA been satisfied for the electronic record as produced? Third, integrity: is there reliable evidence that the offered exhibit is the same object, or a faithful derivative, as the one acquired from the stated source? Fourth, provenance and authenticity: what evidence supports the claim that the depicted person, voice or event is genuine rather than synthetic or materially altered? Fifth, weight: after considering provenance, expert evidence, contextual corroboration and contrary explanations, how much evidentiary weight should the item receive?

This sequence prevents two common mistakes. One is treating section 63 compliance as proof that the content is true. The other is treating any authenticity challenge as an admissibility veto. A properly certified file may still be fake; a genuinely recorded event may still arrive with incomplete provenance; and a synthetic reconstruction may itself be admissible when offered

³⁵ Bharatiya Sakshya Adhiniyam, 2023, s. 39(2).

for a different proposition, such as proving that a defamatory deepfake was circulated. The legal meaning of authenticity is therefore inseparable from the purpose for which the exhibit is tendered.

IX. CONCLUSION

India's 2026 SGI framework is significant because it moves regulation beyond reactive content moderation. It introduces user declarations, platform verification, visible labelling, permanent metadata and technical provenance. Those measures are usually discussed as tools for online transparency. They are also potential evidence. Once a dispute reaches court, however, their value depends on whether the information survived the takedown process in a form that can be examined, certified and challenged.

The BSA supplies a strong foundation for electronic-record integrity through section 63, certificates and hashing. Recent Supreme Court authority confirms that statutory electronic-evidence requirements remain consequential. But synthetic media exposes a conceptual gap: preserving an unchanged file is not the same as proving that the world depicted in that file was real. Courts need provenance around the file, not merely integrity of the file.

The proposed Preserve-Before-Takedown protocol reconciles the two legal imperatives. Harmful SGI can be removed from public view at regulatory speed while a sealed, proportionate and hash-linked evidence bundle is preserved for adjudication. The bundle should capture native media, context, platform identifiers, user declarations, verification outputs, provenance metadata, moderation events and chain-of-custody information, with controlled access for both sides. This approach does not ask courts to trust AI detectors or platforms as final arbiters of truth. It does the opposite: it preserves the material necessary for human adjudication to remain meaningful in an era when realistic media can no longer be assumed to be real.