

WHEN SEEING IS NO LONGER BELIEVING: DEEPFAKES AND THE CRISIS OF DIGITAL EVIDENCE

By Gayatri Ponnamparambil and Ananya Pathak

Institute of Law, Nirma University

ABSTRACT

The emergence of deepfakes presents a growing challenge for courts that rely on digital evidence to establish contested events. As synthetic media becomes increasingly difficult to distinguish from authentic recordings, the evidentiary problem extends beyond the admission of electronic records to determining whether the material itself corresponds to reality. This article examines the adequacy of India's existing evidentiary and regulatory framework in addressing this uncertainty. It considers the Bharatiya Sakshya Adhiniyam, 2023, judicial safeguards governing electronic evidence, and India's emerging regulation of synthetically generated information. Particular attention is given to the "liar's dividend", through which the existence of convincing deepfakes may be used to cast doubt on genuine evidence. The article asserts that formal certification and platform labelling cannot independently settle disputed authenticity. It suggests a multi-layered evidentiary strategy that integrates provenance, forensic analysis, expert evaluation, and support, allowing courts to differentiate advanced forgeries from authentic digital evidence without considering the mere chance of a deepfake as enough reason for skepticism. The article's original contribution is to locate the weakness of the framework in the gap between certifying how a record was produced and establishing that the event it depicts occurred, and to propose a staged judicial protocol that requires a particularised challenge before authenticity is contested.

INTRODUCTION

Imagine a video is presented before a court. It appears to show an event with unmistakable clarity. The opposing party does not dispute that the video exists. Instead, it responds with a deceptively simple assertion: "It is AI-generated." If the footage is genuine, the question then becomes not merely whether the court can admit digital evidence, but whether it can still determine when digital evidence is authentic.

This is the emerging evidentiary problem created by deepfakes. Artificial intelligence can now create or alter faces, voices, images, and videos with greater realism, making fake events seem authentic. The magnitude of the issue is also significant as well. The Identity Fraud Report 2025 recorded one deepfake attempt globally every five minutes in 2024 and reported that deepfakes accounted for 40 per cent of biometric fraud.¹ The problem extends beyond celebrities and political figures whose likenesses have already been exploited through synthetic media; it reaches the increasingly digital foundations of legal proceedings, where CCTV footage, recordings, messages and other electronic records may materially influence the outcome of a case. Indian courts have already begun confronting AI-enabled misuse of identity and likeness, while the evidentiary framework governing electronic records has evolved from the Indian Evidence Act, 1872 to the Bharatiya Sakshya Adhiniyam, 2023.

Yet deepfakes create a paradox more serious than the possibility of false evidence being believed as true. Once convincing fabrication becomes commonplace, genuine evidence may itself become easier to dismiss as fabricated the phenomenon described as the “liar’s dividend.”² The central challenge for law is therefore twofold: preventing synthetic evidence from masquerading as truth while ensuring that genuine evidence is not rejected merely because artificial alternatives have made visual authenticity uncertain.

This article examines whether India's existing evidentiary and regulatory framework can meet that challenge, and argues for a stronger system centred on provenance, forensic verification and corroboration, rather than appearance alone.

The article’s contribution lies in three linked claims. First, the weakness in the present framework lies less in the rules for admitting electronic records than in what they verify: Sections 61 to 63 of the BSA address how a record was produced, stored and certified, but say little about whether the event it depicts took place. Secondly, the liar’s dividend is a defining problem for courts, because a framework that answers fabrication with blanket suspicion merely transfers the harm to genuine evidence. Thirdly, the article proposes a staged judicial protocol that conditions heightened scrutiny on a particularised and credible challenge and separates admissibility, authenticity and evidentiary weight. It is designed to work within the existing statutory structure.

¹Entrust Cybersecurity Institute, 2025 Identity Fraud Report (Entrust, November 2024).

²Bobby Chesney and Danielle Citron, ‘Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security’ (2019) 107 California Law Review 1753.

THE DEEFAKE PROBLEM: WHEN THE ARTIFICIAL BECOMES INDISTINGUISHABLE FROM THE REAL

Deepfakes are a form of synthetic media in which artificial intelligence is used to generate or manipulate audiovisual material, including faces, voices, images and videos. Unlike earlier forms of digital manipulation, contemporary generative systems can produce synthetic content that is increasingly difficult to distinguish from recordings of real people or events.³

The technology has also evolved beyond a single form of manipulation. Face-swapping can transfer someone's appearance onto another individual's body; voice-cloning can mimic a person's speech and vocal traits; and generative models can produce completely artificial scenes instead of simply modifying an existing recording. This advancement is important because the evidentiary worth of audio-visual content has historically relied, at least in part, on the belief that a recording depicts an event that took place in front of a camera or microphone.

Generative AI increasingly challenges the assumption that synthetic media may have no underlying real-world event at all.⁴ The consequences are already visible beyond the courtroom. Public figures have been targeted through fabricated videos, cloned voices and manipulated representations, while deepfake technology has also facilitated reputational harm, impersonation, fraud and non-consensual sexual imagery. The emergence of what has been described as an “authenticity infrastructure” reflects growing efforts to establish media provenance, integrity and manipulation across the journey from recording to distribution.⁵

In India, the 2023 deepfake involving actor Rashmika Mandanna became a prominent illustration of how a person's face could be convincingly superimposed onto another individual's body and rapidly circulated online.⁶ The incident brought the accessibility and realism of synthetic media into mainstream public attention. Yet the most significant danger for the justice system is not simply that false material can appear genuine. It is that the growing sophistication of synthetic media may make genuine material appear contestable. As the distinction between authentic and fabricated audiovisual content becomes increasingly

³Raphaël Millière, ‘Deep Learning and Synthetic Media’ (2022) 200 *Synthese* 231.

⁴*ibid.*

⁵Sam Gregory, ‘Deepfakes, Misinformation and Disinformation and Authenticity Infrastructure Responses: Impacts on Frontline Witnessing, Distant Witnessing, and Civic Journalism’ (2022) 23(3) *Journalism* 708.

⁶‘Deepfake video of Rashmika Mandanna: How to identify deepfakes’ *The Indian Express* (9 November 2023).

difficult to establish through perception alone, the problem shifts from merely detecting manipulation to determining what evidence can be trusted and why.

FROM FAKE EVIDENCE TO THE “LIAR’S DIVIDEND”

The concern identified above takes on greater significance when deepfakes enter the courtroom, where fabricated material may be mistaken for genuine evidence. A manipulated CCTV recording, cloned voice, altered photograph or synthetic video may be presented before a court as an authentic representation of an event. The difficulty is heightened because increasingly sophisticated deepfakes can be difficult for ordinary viewers to distinguish from genuine material, making appearance an unreliable indicator of authenticity.⁷

The more difficult problem, however, operates in the opposite direction. The existence of convincing deepfakes can itself challenge genuine digital evidence, undermining its credibility and enabling authentic footage to be dismissed as fabricated.⁸ This phenomenon, known as the “liar’s dividend”, arises when an individual confronted with authentic evidence exploits the existence of synthetic media by claiming that the evidence is AI-generated. The mere possibility that convincing audiovisual material can be fabricated may therefore create sufficient uncertainty to undermine the credibility of genuine evidence.

The phenomenon does not mean that every allegation of a deepfake will automatically succeed. Its significance lies instead in the new form of uncertainty it introduces. Research involving more than 15,000 participants has demonstrated that false claims that information was fabricated can generate a reputational advantage.⁹ This distinction is important because it prevents the liar’s dividend from being treated as an inevitable consequence of every deepfake allegation. Rather, its danger lies in providing a strategic means of exploiting uncertainty where the authenticity of digital material is already difficult to establish.

Human perception cannot reliably resolve this uncertainty. Experimental research has demonstrated that people can struggle to distinguish manipulated audiovisual material from authentic recordings, while human and machine detection systems can also make different

⁷Matthew Groh and others, ‘Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds’ (2022) 119(1) Proceedings of the National Academy of Sciences e2110013119.

⁸Sam Gregory, ‘Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism’ (2022) 23(3) Journalism 708.

⁹Kaylyn Jackson Schiff, Daniel S Schiff and Natália S Bueno, ‘The Liar’s Dividend: Can Politicians Claim Misinformation to Evade Accountability?’ (2025) 119(1) American Political Science Review 71.

types of errors.¹⁰The concept of an “Indistinguishability Threshold” captures the point at which human ability to distinguish synthetic content from authentic material approaches random chance. Once this threshold is crossed, relying on the content itself to determine authenticity becomes increasingly unreliable, making provenance and other forms of external verification more important.¹¹

The consequences are therefore potentially symmetrical. Acceptance of fabricated evidence may contribute to the wrongful conviction of an innocent person, while rejection of genuine evidence may allow a culpable person to escape responsibility. The appropriate legal response cannot consequently be to treat all digital evidence with suspicion. If the existence of deepfakes becomes sufficient to undermine every contested recording, the technology could damage the credibility of genuine evidence even where no manipulation has occurred.¹²

The challenge is therefore no longer merely to identify what is false, but to establish what can reliably be believed and why. This shifts the evidentiary question from appearance to authentication and provides the foundation for examining whether India's existing legal framework is capable of resolving disputes over the authenticity of digital evidence.

CAN INDIAN EVIDENCE LAW KEEP UP?

India's evidentiary framework has already adapted to the increasing importance of electronic records. Under the Indian Evidence Act, 1872, Sections 65A and 65B created a specific framework for proving electronic records.¹³ The Bharatiya Sakshya Adhiniyam, 2023 (‘BSA’) retains this approach in revised form. Section 61 recognises electronic or digital records as having the same legal effect as other documents, subject to Section 63, while Section 62 provides that their contents are to be proved in accordance with Section 63.¹⁴

Section 63 is particularly significant. It provides for the admissibility of electronic records where specified conditions concerning the computer or communication device and the production of the record are satisfied. Section 63(4) further requires a certificate identifying

¹⁰Groh and others (n 1).

¹¹Alexander Loth, ‘The Indistinguishability Threshold: Measuring Cognitive Vulnerabilities to AI-Generated Disinformation’ in 18th ACM Web Science Conference (WebSci Companion ’26) (ACM 2026).

¹²Sam Gregory, ‘Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism’ (2022) 23(3) Journalism 708.

¹³Indian Evidence Act 1872, ss 65A–65B.

¹⁴Bharatiya Sakshya Adhiniyam 2023, ss 61–63.

the electronic record, describing its production and providing relevant particulars of the device involved. The certificate must also be accompanied by the signatures of the person responsible and an expert.¹⁵ The framework therefore seeks to establish the reliability and manner of production of an electronic record before admission.

Yet certification does not necessarily resolve the deeper problem created by deepfakes. A certificate may establish the manner in which a file was produced or preserved without establishing that the event depicted in it actually occurred. A perfectly preserved deepfake remains a deepfake.¹⁶ Similarly, a hash value can demonstrate that a particular file has remained unchanged since hashing, but cannot by itself establish that the file was authentic when created.¹⁷ The distinction becomes pertinent where the opposing party accepts the file's existence but challenges whether its contents represent reality.

The limits of this framework become clearer when Sections 61 to 63 are read in sequence. Section 61 is essentially a rule of non-discrimination: an electronic record cannot be denied admissibility merely because of its form and, subject to Section 63, has the same legal effect as any other document. It speaks to form, not truth. A deepfake is formally an electronic record; Section 61 gives it no advantage over a genuine recording, but supplies no criterion for telling the two apart. Section 62 is similarly procedural, channelling proof of electronic content through Section 63 without prescribing what the court must be satisfied of afterwards.¹⁸

Section 63 therefore carries the weight, and its limitations are structural. The conditions in sub-section (2) concern the device: regular use in the ordinary course of activities, regular feeding of relevant information, proper operation during the material period, and a record that reproduces information so fed.¹⁹ These conditions test the reliability of the machine, not of the content. A synthetic video produced on a properly functioning, lawfully controlled device may satisfy each of them. They presuppose a record that captures something that happened, not one generated without any underlying event.

¹⁵ibid s 63(4).

¹⁶Aishwarya Pandit & Vaishnavi Shetty, Digital Deception: AI and the Law of Evidence, 40 SUPREMO AMICUS [1] (2026).

¹⁷Sunita Meena, 'When Machines Testify: Artificial Intelligence and the Bharatiya Sakshya Adhiniyam, 2023' (2026) 7(8) International Journal of All Research Writings 78.

¹⁸Bharatiya Sakshya Adhiniyam 2023, ss 61–62.

¹⁹ibid s 63(2).

The Section 63(4) certificate compounds this. It identifies the record, describes its production, gives particulars of the device and is signed by the person in charge and an expert, with the hash value disclosed in the prescribed form.²⁰ Each element serves the integrity of the file as produced and preserved. A hash value fixes a file's state at the moment it is computed; it cannot show that the file was authentic when created, and an expert's certification attests to technical particulars rather than to whether the footage depicts a real event. Moreover, sub-section (1) renders output admissible without production of the original, so the court may be denied the metadata and device-level artefacts on which forensic detection often depends. Finally, once a record is admitted, the BSA does not direct how a court should treat a specific allegation that it is AI-generated, what each party must produce, or what degree of satisfaction is required. Those questions are left to general principles of relevance, burden of proof and appreciation of evidence.²¹

Indian jurisprudence has long recognised the need for safeguards against manipulation of recorded evidence. In *Ram Singh v Col Ram Singh*, the Supreme Court emphasised identification of the speaker's voice, proof of accuracy, exclusion of tampering or erasure, and safe custody of the recording. The Court stressed that recorded evidence must be received with caution and subjected to rigorous verification, including scrutiny of its accuracy and possibility of manipulation. In the context of modern digital evidence, this can extend to indicators such as timestamps, device identifiers, server logs and other metadata.²²

The Court subsequently strengthened these safeguards in *Anvar PV v PK Basheer*, where it emphasised compliance with the statutory requirements governing electronic records, and in *Arjun Panditrao Khotkar v Kailash Kushanrao Gorantyal*, where it reaffirmed the importance of the certificate requirement for the admissibility of electronic evidence.²³ These principles remain valuable, but they developed in a context where the principal concern was whether an existing recording had been altered.

Indian courts have long treated authentication as a condition of reliance on recorded material. In *Yusufalli Esmail Nagree v State of Maharashtra* the Supreme Court treated a contemporaneous tape-recording as admissible in the absence of tampering, and in *RM*

²⁰ibid s 63(4) and the Schedule (Parts A and B).

²¹ibid s 63(1), (4).

²²*Ram Singh v Col Ram Singh* (1985) 3 SCC 611.

²³*Anvar PV v PK Basheer* (2014) 10 SCC 473; *Arjun Panditrao Khotkar v Kailash Kushanrao Gorantyal* (2020) 7 SCC 1.

Malkani v State of Maharashtra it required relevance, identification of the voice, and proof of accuracy by eliminating the possibility of erasure.²⁴ *Ram Singh*, discussed above, consolidated these requirements.

The electronic-records decisions shifted the emphasis to the method of proof. *Anvar* held that electronic records, being susceptible to tampering and alteration, must be proved through the statutory safeguards, which it described as directed at ensuring source and authenticity.²⁵ *Shafhi Mohammad v State of Himachal Pradesh* suggested that the certificate could be relaxed where a party could not produce it,²⁶ but *Arjun Panditrao* overruled that view, restored *Anvar*, and held the certificate necessary for secondary electronic evidence, though not where the original device itself is produced.²⁷ Indian law has thus consistently insisted on source and authenticity, but pursued them through proxies such as voice identification, accuracy, custody and certified origin, designed for a setting in which the fear was alteration of an existing recording.

The Supreme Court has now acknowledged the wider problem. In *Pune Bar Association v Union of India*, a three-Judge Bench upheld the validity of Section 63(4) and its Schedule, describing the hash value as an electronic fingerprint, treating expert certification as an additional layer of authenticity, and recognising that electronic records are vulnerable to alteration, a concern heightened by artificial intelligence and deepfake technology.²⁸ It also indicated that a person with special skill in computer science and cyber forensics may, if the court is satisfied on unimpeachable material, sign Part B of the Schedule, leaving the wider scope of expert certification open.²⁹ The decision validates the certification regime, but does not tell a trial court how to resolve a contested allegation that a certified record is synthetic. That must still proceed through provenance, forensic opinion and corroboration.

Your One Stop Legal Destination

The BSA also provides for technical expertise. Section 39 recognises expert opinion on scientific and specialised matters, including the opinion of an Examiner of Electronic Evidence under Section 79A of the Information Technology Act, 2000.³⁰ Yet expert

²⁴*Yusufalli Esmail Nagree v State of Maharashtra* AIR 1968 SC 147; *RM Malkani v State of Maharashtra* (1973) 1 SCC 471.

²⁵*Anvar PV v PK Basheer* (2014) 10 SCC 473.

²⁶*Shafhi Mohammad v State of Himachal Pradesh* (2018) 2 SCC 801.

²⁷*Arjun Panditrao Khotkar v Kailash Kushanrao Gorantyal* (2020) 7 SCC 1.

²⁸*Pune Bar Association v Union of India* WP(C) 599/2026, decided on 22 May 2026 (SC).

²⁹*ibid.*

³⁰*Bharatiya Sakshya Adhiniyam* 2023, s 39(2); *Information Technology Act* 2000, s 79A.

examination cannot necessarily produce an infallible binary answer where sophisticated synthetic media can evade detection and detection systems themselves operate with varying accuracy.

The resulting gap is therefore not that Indian law cannot admit or examine digital evidence. The underlying issue, however, is the fact that the emergence of generative AI poses an issue that the electronic-evidence paradigm was not built to answer does the digital evidence reflect any actual physical world occurrence?

DISTINGUISHING ADMISSIBILITY, AUTHENTICITY, INTEGRITY, PROVENANCE, RELIABILITY AND EVIDENTIARY WEIGHT

These terms are often used interchangeably, although each answers a different question, and a record can succeed on one and fail on another.

Admissibility is the threshold question of whether the court may receive the record, turning on relevance and compliance with Sections 61 to 63.³¹ Clearing it decides nothing about truth. Authenticity asks whether the record is what its proponent claims: that it depicts the persons, event, time and place alleged and was not fabricated. Deepfakes place this question in dispute.

Integrity asks whether the record has remained unaltered since a given moment, such as capture, seizure or hashing. It is necessary for authenticity but does not establish it, since a record intact from seizure may have been fabricated before then. Provenance is the documented origin and history of the record, including its creator, device, edits and chain of transfer. It bears on both authenticity and integrity but is distinct from each: its absence does not prove fabrication, and its presence does not prove the depicted event was real.

Reliability concerns whether a process can be trusted to produce accurate results, whether the device that produced the record, the concern of Section 63(2), or the tools used to examine it, including detection software whose error rates may be unknown. Evidentiary weight is the probative value finally assigned to admitted material after considering all the evidence; a record may be admissible and authentic yet carry limited weight because it is partial, poor in quality or ambiguous.

³¹Bharatiya Sakshya Adhiniyam 2023, ss 61–63.

Keeping these apart guards against two errors. Treating certification as proof of authenticity risks admitting a perfectly preserved fabrication. Treating the bare possibility of fabrication as a ground for exclusion confuses doubt about authenticity with a defect in admissibility and hands the liar's dividend to the party who benefits from it. The protocol proposed below rests on this separation.

INDIA'S EMERGING LEGAL RESPONSE

Indian courts have not remained entirely silent as synthetic media has moved from a technological novelty to a practical source of harm. Their response, however, has largely developed through existing doctrines of personality rights, privacy, reputation and intermediary liability rather than through a dedicated legal regime for deepfakes.

Indian courts have already begun adapting existing legal principles to AI-enabled synthetic media. In *Anil Kapoor v Simply Life India*, the Delhi High Court addressed the unauthorised use of the actor's name, image, voice and likeness through technologies including AI and deepfakes, demonstrating that courts have had to confront representations of individuals that may never have existed in reality.³² Additionally, during the 2024 Lok Sabha elections, AI-generated voice cloning and manipulated audiovisual content were reportedly used on a large scale, including an AI-manipulated video falsely portraying Amit Shah as supporting changes to India's reservation system.³³ These developments demonstrate how synthetic media can manufacture statements and events that appear authentic, making the reliability of audiovisual information increasingly difficult to assess.

The judicial response has continued to develop. In *Raghav Chadha v Ashok Kumar John Doe*, the Delhi High Court considered AI-generated deepfakes and morphed content, demonstrating that Indian courts are already confronting the evidentiary implications of synthetic media.³⁴ The significance of such cases for the present discussion lies not merely in

³²*Anil Kapoor v Simply Life India & Ors* CS(COMM) 652/2023, orders dated 20 September 2023 and 30 April 2024 (Delhi HC).

³³Eshaan Mani, 'Shaping Robust AI Regulation: Lessons from India's "Deepfake" Election' *Harvard Political Review* (6 December 2024).

³⁴*Raghav Chadha v Ashok Kumar John Doe & Ors* CS(OS) 466/2026, judgment dated 1 July 2026 (Delhi HC).

the harm caused by fabricated content, but in the growing difficulty of determining whether digital material placed before a court is genuine or artificially generated.³⁵

Moreover, India's regulatory response has moved beyond judicial innovation. On 10 February 2026, MeitY notified the *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026* through GSR 120(E), which came into force on 20 February 2026.³⁶ The amendments expressly introduce “synthetically generated information” (SGI) into the intermediary framework and impose additional obligations concerning such content. These include user disclosures, prominent labelling of SGI and, where technically feasible, metadata or other technical measures capable of identifying its provenance.³⁷

This represents an important shift: Indian regulation now expressly recognises synthetic media rather than addressing deepfakes only through pre-existing offences and rights. Yet the framework primarily regulates the creation and dissemination of synthetic content through intermediaries. It does not, by itself, answer the courtroom question at the centre of this article. India's emerging regulatory framework is therefore an important step, but it does not eliminate the evidentiary uncertainty created when synthetic and genuine media become increasingly difficult to distinguish.

THE REGULATORY GAP: LABELLING IS NOT AUTHENTICATION

India's emerging regulatory framework represents an important step towards addressing synthetic media, but it does not fully resolve the evidentiary problem identified above. The 2026 IT Rules require certain synthetically generated information to be identified and labelled, alongside technical measures such as metadata and provenance information.³⁸ Yet labelling answers a different question from the one a court must answer. A platform may identify whether content is synthetic whereas, a court must determine whether a digital record is authentic and whether the event it depicts actually occurred.

³⁵Aishwarya Pandit & Vaishnavi Shetty, Digital Deception: AI and the Law of Evidence, 40 SUPREMO AMICUS (2026).

³⁶*Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2026*, GSR 120(E), Gazette of India, 10 February 2026.

³⁷Ministry of Electronics and Information Technology, ‘Frequently Asked Questions: Amendments to IT Rules, 2021 relating to SGI’ (10 February 2026).

³⁸*Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2026*, GSR 120(E), Gazette of India, 10 February 2026.

This distinction becomes particularly important where genuine evidence is challenged as AI-generated. Provenance can establish the origin and history of a digital asset, but provenance alone cannot establish that the content itself is factually true.³⁹ C2PA Content Credentials, for instance, can create a cryptographically verifiable record of an asset's creation and modifications, but cannot determine whether the underlying event depicted is true.⁴⁰

This points towards the need for an “authenticity infrastructure”: systems capable of tracing media provenance, integrity and manipulation from capture through editing and distribution, rather than relying solely on the appearance of the final file.⁴¹ Such mechanisms could provide courts with information about a recording's source and modification history, supplementing forensic examination and judicial assessment.

However, provenance mechanisms are not infallible. Content Credentials may be absent or incomplete, and a valid provenance record does not necessarily establish the truth of what a recording depicts.⁴² This is particularly significant for older CCTV footage, screenshots and recordings transferred across platforms, where the original provenance chain may no longer exist. The regulatory gap is therefore not simply the absence of rules governing deepfakes. India is increasingly regulating synthetic content, but it has yet to establish a comprehensive evidentiary framework for resolving contested authenticity where genuine recordings and sophisticated deepfakes may appear equally credible.

TOWARDS A RELIABLE STANDARD OF DIGITAL PROOF

The solution cannot be to treat all digital evidence with suspicion, nor can authenticity be determined by a single AI-detection tool. A reliable standard of digital proof should instead combine provenance, forensic examination, expert evidence and corroboration. First, provenance should form an important part of authentication.⁴³ Where available, courts should examine the source of a recording, the device that captured it, its creation history, subsequent modifications and its chain of transfer. Emerging provenance technologies can provide

³⁹Coalition for Content Provenance and Authenticity (C2PA), ‘C2PA and Content Credentials Explainer’ (2026).

⁴⁰Ibid.

⁴¹Sam Gregory, ‘Deepfakes, Misinformation and Disinformation and Authenticity Infrastructure Responses: Impacts on Frontline Witnessing, Distant Witnessing, and Civic Journalism’ (2022) 23(3) *Journalism* 708, doi:10.1177/14648849211060644.

⁴²C2PA, ‘Frequently Asked Questions: Content Credentials’ (2026).

⁴³Coalition for Content Provenance and Authenticity (C2PA), ‘C2PA and Content Credentials Explainer’ (2026).

cryptographically verifiable records of this history, although they should supplement rather than replace judicial scrutiny.⁴⁴

Second, forensic examination should be used where authenticity is genuinely disputed. Metadata, file structure, compression patterns, device information and other technical indicators may assist in identifying manipulation, although courts should remain cautious about treating detection systems as infallible where their conclusions are probabilistic. Third, expert evidence should assist rather than replace judicial assessment.

Finally, corroboration should remain important. CCTV from another camera, witness testimony, timestamps, physical evidence or other independent records may strengthen or undermine the authenticity of disputed material. The objective should therefore not be to create an impossible test of absolute certainty, but a structured, multi-layered process capable of distinguishing genuine evidence from sophisticated fabrication while preventing the mere allegation of a deepfake from discrediting authentic evidence.

A STEP-BY-STEP JUDICIAL PROTOCOL FOR CONTESTED DIGITAL EVIDENCE

These elements are useful only if arranged into a procedure a trial court can follow. The protocol below operates within the existing framework of the BSA and is engaged only where authenticity is genuinely in dispute, so that the cost and intrusiveness of the inquiry rise only with the strength of the challenge.

The inquiry begins with the statutory gateway. The court decides relevance and compliance with Sections 61 to 63, including the Section 63(4) certificate and hash value.⁴⁵ Non-compliance ends the matter; compliance means only that the record may be considered, and the court should say expressly that authenticity remains undecided.

Heightened scrutiny should then depend on a particularised challenge. A party alleging that the record is synthetic or altered must identify what is said to be fabricated and on what material, such as anomalies in the content, inconsistent metadata, a different original or an expert report. A bare assertion that the material is AI-generated should not oblige the court to

⁴⁴ibid.

⁴⁵Bharatiya Sakshya Adhiniyam 2023, ss 61–63, and the Schedule (Parts A and B).

order forensic examination or shift any burden, since otherwise the liar's dividend would be available on demand. The proponent keeps the burden of showing the record is what it claims to be; the objector need only raise a specific and credible doubt, not prove fabrication. Absent such a challenge, the court proceeds on the proof already given, retaining power to inquire further where circumstances warrant.

Where the challenge crosses that threshold, the court turns first to provenance. It directs the proponent to disclose the device or account used for capture, the person who captured the record, the circumstances, the chain of transfer, whether an original or native file exists, and any available provenance information such as content credentials.⁴⁶ The hash value recorded at the earliest practicable point should be compared with that of the tendered file, and unexplained refusal to produce an original may be weighed against the refusing party.

If doubt survives, the court should obtain forensic examination from an Examiner of Electronic Evidence or another person whose expertise it is satisfied is adequate.⁴⁷ The expert should disclose methodology, tools and known error rates, and state conclusions as degrees of support, such as consistent with authentic capture, showing indicators of manipulation, or inconclusive, rather than as a binary verdict. Detection output is one input to the court's assessment, not a substitute for it, and each side may cross-examine the expert and call its own. The court then tests these findings against independent corroboration, such as other cameras, witnesses, device and network logs, third-party records and the timeline.

Finally, the court records separate, reasoned conclusions on admissibility, authenticity and weight. On authenticity it should find that the record is established as authentic, that authenticity is not established, or that it is shown to be fabricated. An authentic record is relied on according to its weight. Where authenticity is not established, the court should not rely on the record as the sole basis for a finding, particularly one adverse to an accused. Where fabrication is found, the record is rejected and the tendering party may face the consequences under the law on fabrication of evidence; a challenge raised without any basis may likewise count against the party's credibility and in costs.

⁴⁶Coalition for Content Provenance and Authenticity (C2PA), 'C2PA and Content Credentials Explainer' (2026).

⁴⁷Bharatiya Sakshya Adhiniyam 2023, s 39(2); Information Technology Act 2000, s 79A; *Pune Bar Association v Union of India* WP(C) 599/2026, decided on 22 May 2026 (SC).

CONCLUSION

Deepfakes create an evidentiary dilemma: the court may either place confidence in material that was never real or reject material that was. The latter risk is particularly significant because the liar's dividend allows the existence of deepfakes themselves to generate doubt around authentic material.

India has begun responding through judicial protection of identity and reputation, regulation of synthetically generated information and an evolving framework for electronic evidence. Yet these measures do not fully address the courtroom problem of contested authenticity. The answer cannot be to distrust digital evidence, because that may allow genuine evidence to be discarded; nor can it be to trust digital evidence merely because it satisfies formal requirements.

A reliable evidentiary framework must instead place provenance, forensic examination, expert assessment and corroboration at its centre. In an age when seeing is no longer necessarily believing, justice must depend not merely on what evidence appears to show, but on whether its authenticity can be demonstrated.

The argument leads to three conclusions. First, compliance with Section 63 should be the beginning of the authenticity inquiry, never its end: a certificate answers whether a record may be admitted, not whether the event it shows occurred. Secondly, a bare claim that evidence is AI-generated should carry no evidential force; the party seeking doubt must say, with particulars, why, and only then should the burden of provenance and forensic verification be engaged. Thirdly, once that threshold is crossed, courts should follow the staged protocol proposed above and record separate findings on admissibility, authenticity and weight.

None of this requires fresh legislation. Trial courts can adopt these steps under existing powers, and High Courts can consolidate them through practice directions, building on the Supreme Court's recent engagement with Section 63(4). Reform can then support the courts: rule-makers should require preservation of native files and metadata at seizure, labelling and provenance standards such as the 2026 IT Rules and C2PA should be developed with evidentiary use in mind while recognising that labelling is not authentication, and forensic

capacity, including accredited laboratories and published error rates for detection tools, must expand.

What is at stake is the integrity of fact-finding in both directions. A system that accepts every convincing recording will convict the innocent; one that doubts every recording will acquit the guilty. The task is to build procedures that ask the right question at the right stage, and India's courts are now in a position to begin.



Your One Stop Legal Destination