WHEN REALITY BECOMES EVIDENCE: DEEPFAKES, CRIMINAL ATTRIBUTION AND DIGITAL AUTHENTICITY IN INDIA

Angeri Thyagaraju, LL.B., Aurora Higher Education and Research Academy
(Deemed-to-be University), Bhongir Campus

ABSTRACT

The rapid development of generative artificial intelligence has fundamentally altered the evidentiary value traditionally attached to photographs, videos and audio recordings. Deepfakes can manufacture apparently authentic representations of persons, events and statements that never occurred. The resulting legal problem extends beyond misinformation or privacy: it challenges the ability of criminal law to establish what actually happened, whether the digital record is authentic, and who should be held criminally responsible for its creation or dissemination. This article argues that India's existing legal framework contains several provisions capable of addressing deepfake-related conduct, but remains fragmented when confronted with the distinctive problems of digital authenticity and criminal attribution. The Bharatiya Sakshya Adhiniyam, 2023 recognises electronic and digital records as evidence; the Bharatiya Nyaya Sanhita, 2023 addresses false electronic records and forgery; and the Information Technology Act, 2000 provides offences that may apply to identity theft, personation, privacy violations and specified forms of unlawful electronic publication. India's 2026 amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 further introduce a dedicated due-diligence framework for synthetically generated information. Yet admissibility of an electronic record does not establish the truth of the event represented by that record, nor does it identify the person responsible for its creation. This article proposes a four-layer approach to digital authenticity—file integrity, source authenticity, event authenticity and criminal attribution—as a framework for judicial assessment of deepfake evidence. It contends that India's response should focus not merely on regulating synthetic content, but on developing reliable methods for separating the existence of a digital file from the authenticity of the reality it purports to depict.

Keywords: Deepfakes, Artificial Intelligence, Digital Evidence, Criminal Attribution, Electronic Records, Digital Authenticity, Bharatiya Sakshya Adhiniyam, Bharatiya Nyaya Sanhita.

I. INTRODUCTION: WHEN APPEARANCE NO LONGER PROVES REALITY

For much of the history of evidence, photographs, audio recordings and videos possessed a powerful intuitive quality: they appeared to show what had actually happened. Their evidentiary value was never absolute, but their capacity to reproduce an event gave them a special relationship with factual reality. Generative artificial intelligence has disrupted that assumption.

A deepfake can place a person's face on another person's body, reproduce a person's voice, manufacture apparently authentic statements, or create an event that never occurred. The problem is therefore not simply that false information can circulate more rapidly. The deeper problem is that the appearance of evidence can now be manufactured independently of the event it purports to record.

This creates a difficult question for criminal law: if a video apparently shows an accused person committing an offence, what exactly must the prosecution prove? It is insufficient merely to establish that a video file exists. It may also be insufficient to demonstrate that the file has remained technically unaltered after collection. A perfectly preserved deepfake remains a deepfake.

India's legal system addresses this problem through several overlapping regimes. The Bharatiya Sakshya Adhiniyam, 2023 (BSA) recognises electronic and digital records and provides a statutory framework governing their proof and admissibility.[1] The Bharatiya Nyaya Sanhita, 2023 (BNS) addresses false documents and false electronic records, including forgery.[2] The Information Technology Act, 2000 (IT Act) separately addresses identity theft, cheating by personation, privacy violations and specified forms of unlawful electronic publication.[3] More recently, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 introduced specific intermediary due-diligence requirements concerning synthetically generated information (SGI).[4]

These provisions are important, but they do not by themselves resolve the central evidentiary problem: how should a criminal court distinguish an authentic digital record from an authentic-looking fabrication, and how should it attribute the fabrication to a particular person?

This article argues that India's deepfake problem should therefore be understood principally as

a problem of digital authenticity and criminal attribution, rather than merely as a problem of content regulation.

II. DEEPFAKES AND THE COLLAPSE OF DIGITAL AUTHENTICITY

The legal difficulty created by a deepfake arises from the distinction between a digital object and the reality represented by that object.

Consider a hypothetical video showing A threatening B. If the prosecution produces the video, several different propositions may be involved: (1) a particular electronic file exists; (2) the file was obtained from a particular device or platform; (3) the file has not been altered after collection; (4) the images or sounds contained in the file correspond to an actual event; (5) A is genuinely the person depicted; (6) A actually made the alleged statement; and (7) A was responsible for creating or disseminating the recording.

These propositions are not interchangeable.

The BSA's treatment of electronic records is therefore only the beginning of the evidentiary inquiry. Section 61 recognises electronic or digital records as records for evidentiary purposes; section 62 contains special provisions concerning evidence relating to electronic records; and section 63 establishes the conditions under which specified electronic records are admissible.[5]

The important distinction is this: admissibility concerns whether evidence may be received by a court; authenticity concerns what confidence the court should place in the proposition the evidence is offered to establish.

A deepfake exposes the limits of treating these questions as identical.

Traditional electronic-evidence doctrine has concentrated substantially on the conditions governing the proof and admissibility of electronic records. The Supreme Court's jurisprudence under section 65B of the Indian Evidence Act, 1872, particularly in Anvar P.V. v. P.K. Basheer and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, illustrates the importance of statutory requirements for electronic evidence.[6] The present BSA retains a specialised statutory framework for electronic records.[7]

Yet a deepfake can satisfy technical requirements concerning the electronic record while

remaining factually false.

The challenge is therefore no longer simply: Is this electronic record genuine? It must become: What, precisely, is genuine about this electronic record?

III. CRIMINAL ATTRIBUTION: WHO IS RESPONSIBLE FOR THE DEEPFAKE?

Deepfakes create an attribution chain that conventional offences do not always encounter with the same intensity.

A synthetic video may pass through several actors: Creator → Manipulator → First Uploader → Platform → Forwarder → Viewer.

The legal responsibility of these actors cannot automatically be treated as identical.

The person who generates a deepfake with criminal intent is conceptually different from a person who unknowingly forwards it. Similarly, a person who knowingly uploads manipulated content for the purpose of deceiving a victim may occupy a different legal position from an intermediary that merely hosts the content.

This distinction is particularly important because criminal liability ordinarily depends upon both conduct and the legally required mental element. The existence of harmful content cannot, without more, establish the identity or intention of the person who created it.

The BNS provides a particularly relevant framework. Section 335 defines circumstances in which a person makes a false document or false electronic record, including dishonest or fraudulent creation or unauthorised alteration of an electronic record.[8] Section 336 then separately criminalises forgery and expressly applies where a person makes a false document or false electronic record with specified fraudulent, injurious or reputational purposes.[9] The latter provision is especially relevant to deepfakes because a fabricated digital representation may be created precisely to cause injury, facilitate cheating, commit fraud or damage reputation.

The IT Act also provides potentially relevant offences. Section 66C addresses identity theft, section 66D addresses cheating by personation using a computer resource, and section 66E addresses violation of privacy. Sections 67 and 67A address specified categories of obscene

and sexually explicit electronic material.[1] These provisions may apply to particular deepfake[0] conduct, but their application depends on the facts and the elements of the individual offence.

However, the existence of potentially applicable offences does not eliminate the attribution problem.

Suppose a manipulated video is uploaded from an anonymous account. Investigators may identify an IP address, a device, an email account, a telephone number or a platform account. None of these identifiers necessarily establishes, by itself, that the particular accused person created the deepfake. Devices can be shared. Accounts can be compromised. Credentials can be stolen. Networks can be masked. Content can be generated by one person and uploaded by another.

The prosecution must therefore establish a chain: digital content → digital source → identified account/device → identified individual → knowledge/intent → prohibited conduct.

A weakness at any link may create reasonable doubt.

This is the attribution gap: the law may recognise the wrongfulness of the deepfake, yet the evidentiary system may struggle to connect the wrong to the person before the court.

IV. THE EVIDENTIARY PROBLEM: FILE INTEGRITY IS NOT EVENT AUTHENTICITY

The most important conceptual distinction in deepfake litigation is between integrity and authenticity.

File integrity asks whether the digital file presented to the court has been altered after it was collected or preserved.

Event authenticity asks whether the content of that file accurately represents an event that actually occurred.

These are different questions.

A hash value may demonstrate that a particular file has remained unchanged since it was hashed. A forensic image may demonstrate that data was extracted from a particular device.

Metadata may provide information about creation, modification or transmission. A certificate may satisfy a statutory evidentiary requirement. But none of these necessarily proves that the underlying event occurred.

For example, a synthetic video may be generated on 1 September, preserved without modification, hashed on 2 September and produced in court on 3 September. The file may be perfectly intact. The event depicted in it may nevertheless never have occurred.

Thus, the central evidentiary mistake in the deepfake era would be to equate technical integrity with factual truth.

Section 63 of the BSA provides detailed conditions for the admissibility of specified electronic records, including conditions concerning the regular use and operation of the relevant computer or communication device and a certificate accompanying the electronic record in the circumstances covered by the provision.[11] Compliance with those conditions establishes a statutory route to admissibility; it does not create a general presumption that every representation contained in an electronic record corresponds to a real-world event.

This distinction becomes particularly important in criminal proceedings because the prosecution must establish the accused's guilt to the applicable standard of proof.

Anvar P.V. v. P.K. Basheer treated the former section 65B framework as a special regime for electronic evidence, while Arjun Panditrao Khotkar reaffirmed the significance of compliance with that statutory framework for secondary electronic evidence.[12] The present BSA should be read in its own statutory terms, but the underlying distinction remains instructive: the legal conditions governing the reception of an electronic record do not eliminate the court's separate task of assessing what factual proposition the record actually proves.

Deepfakes require precisely this next analytical step.

Even after a court concludes that a digital record is admissible, it must still determine what factual weight the record deserves.

V. A FOUR-LAYER MODEL OF DIGITAL AUTHENTICITY

A more effective approach would require courts and investigators to separate four questions.

Layer One: File Integrity. The first question is whether the electronic record produced before the court is the same record that was originally collected. This involves familiar forensic safeguards such as preservation, hashing, chain of custody, device imaging and appropriate documentation. This layer answers: Has the evidentiary file been altered after collection? It does not answer whether the content itself is genuine.

Layer Two: Source Authenticity. The second question concerns origin. Where did the file come from? Was it extracted from a particular device? Downloaded from a social-media platform? Received through a messaging service? Generated by an artificial-intelligence application? Uploaded through an identified account? Source authentication helps establish provenance, but provenance still does not establish truth.

Layer Three: Event Authenticity. The third question is the most difficult. Does the recording accurately depict a real event? Here, courts may need to consider forensic examination of audio and visual characteristics, inconsistencies in metadata, evidence of synthetic generation, contextual evidence, contemporaneous records, independent witnesses and other corroborating material. The important principle is that a video should not be treated as self-proving merely because it looks realistic.

Layer Four: Criminal Attribution. The final question is: Who created, altered, uploaded or disseminated the content, and what did that person know or intend? This requires the digital evidence to be connected to an identifiable individual through a sufficiently reliable chain.

The four layers can therefore be expressed as: Integrity → Provenance → Reality → Attribution.

The framework has a practical advantage: it prevents a weakness at one stage from being disguised by strength at another. A perfectly preserved file cannot compensate for an absence of proof that the event actually occurred. Likewise, proof that a particular person operated an account does not necessarily establish that the person created the synthetic content uploaded through it.

VI. WHY A DEEPFAKE-SPECIFIC OFFENCE ALONE IS NOT ENOUGH

A natural response to deepfakes is to demand a new criminal offence specifically prohibiting their creation or dissemination. Such an offence may be useful in particular circumstances, but

criminalisation alone does not solve the evidentiary problem.

A new offence still requires proof of the existence and nature of the synthetic content; the accused's connection with its creation or dissemination; the required mental element; the relevant prohibited consequence or circumstance; and the reliability of the digital evidence establishing those facts.

The law therefore needs better evidentiary architecture, not merely more offences.

This is particularly important because not every synthetic representation is unlawful. AIgenerated content may be used for satire, education, research, artistic expression, accessibility or legitimate experimentation. An excessively broad criminal prohibition could therefore create significant problems of overcriminalisation and freedom of expression.

The more defensible approach is to regulate harmful conduct and culpable use, while strengthening mechanisms for establishing authenticity and attribution.

VII. TOWARDS A MORE RELIABLE INDIAN APPROACH

India does not necessarily require an entirely separate evidentiary code for deepfakes. Instead, the existing framework should be supplemented by clearer judicial and investigative practices, while the new SGI framework is treated as a regulatory complement rather than a substitute for criminal proof.

First, digital evidence should be accompanied by provenance information wherever reasonably available. Investigators should preserve the circumstances in which the file was obtained rather than merely producing the final copy.

Second, forensic examination should distinguish between file alteration and synthetic generation. A file may have remained untouched after acquisition even though it was artificially generated before acquisition.

Third, courts should seek corroboration where the content itself is reasonably capable of synthetic generation. The greater the capacity of technology to fabricate an event convincingly, the weaker the justification for treating visual realism as independent proof of occurrence.

Fourth, attribution should be treated as an independent evidentiary inquiry. Identification of an

account, device or IP address should not automatically be equated with identification of the human actor.

Fifth, courts should maintain a careful distinction between admissibility and probative value. Admitting a digital record into evidence should not end the inquiry into whether the content accurately represents reality.

The 2026 amendments to the IT Rules strengthen the regulatory layer surrounding synthetic content. MeitY's official materials describe SGI-specific due-diligence obligations, including measures concerning unlawful SGI, labelling or metadata for permissible SGI, user awareness, and additional obligations for Significant Social Media Intermediaries.[13] These measures are significant because they address the circulation and platform-level governance of synthetic content. They do not, however, determine whether a particular deepfake proves a particular fact in a criminal trial or whether a particular accused created it.

Accordingly, intermediary regulation and evidentiary law should be treated as complementary. Platform-level duties may improve provenance, labelling and removal mechanisms, but the judicial question remains one of proof: whether the evidence before the court reliably establishes the fact in issue and connects the relevant conduct to the accused.

VIII. CONCLUSION: FROM TRUSTING THE IMAGE TO TESTING THE CHAIN

Deepfakes expose a foundational weakness in assumptions that have traditionally surrounded audiovisual evidence. The problem is not that courts suddenly have no evidence. The problem is that courts may have too much apparently convincing evidence whose relationship with reality cannot be assumed.

India's contemporary legal framework is not without tools. The BSA recognises electronic and digital records and provides a statutory framework for their admissibility; the BNS addresses false electronic records and forgery; the IT Act contains offences that may apply to identity theft, personation, privacy violations and specified forms of unlawful electronic content; and the 2026 IT Rules introduce additional intermediary obligations concerning synthetically generated information.[14]

The central gap lies elsewhere.

The law must distinguish between the existence of a digital file and the truth of the event depicted by that file. It must also distinguish between the existence of a harmful deepfake and proof of the individual who created or disseminated it.

Accordingly, the deepfake problem should be approached through four distinct questions: Is the file intact? Where did it originate? Does it represent a real event? Who is criminally responsible for it?

These questions correspond to file integrity, source authenticity, event authenticity and criminal attribution.

The significance of this framework extends beyond deepfakes. As synthetic media becomes increasingly indistinguishable from conventional recordings, the criminal justice system cannot continue to rely upon the intuitive proposition that seeing is believing. The evidentiary value of a recording must increasingly depend upon the reliability of the chain connecting data to source, source to event, and event to accused.

The fundamental legal challenge of artificial intelligence is therefore not merely that machines can create false realities. It is that law must determine truth when reality itself can be convincingly manufactured as evidence.

In the deepfake era, the proper question is no longer simply whether the court can see the evidence. It is whether the court can prove what the evidence actually represents.

ENDNOTES

1. Bharatiya Sakshya Adhiniyam, No. 47 of 2023, §§ 61–63 (India).
2. Bharatiya Nyaya Sanhita, No. 45 of 2023, §§ 335–336 (India).
3. Information Technology Act, No. 21 of 2000, §§ 66C–66E, 67, 67A (India).
4. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, Gazette Notification G.S.R. 120(E) (Feb. 10, 2026); Ministry of Electronics & Information Technology, FAQs on Amendments to IT Rules, 2021 Relating to Synthetically Generated Information (Feb. 10, 2026).
5. Bharatiya Sakshya Adhiniyam, No. 47 of 2023, §§ 61–63 (India).
6. Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473; Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.
7. Bharatiya Sakshya Adhiniyam, No. 47 of 2023, §§ 61–63 (India).
8. Bharatiya Nyaya Sanhita, No. 45 of 2023, § 335 (India).
9. Id. § 336.
10. Information Technology Act, No. 21 of 2000, §§ 66C–66E, 67, 67A (India).
11. Bharatiya Sakshya Adhiniyam, No. 47 of 2023, § 63 (India).
12. Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473; Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.
13. Ministry of Electronics & Information Technology, FAQs on Amendments to IT Rules, 2021 Relating to Synthetically Generated Information (Feb. 10, 2026); Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026.
14. Bharatiya Sakshya Adhiniyam, No. 47 of 2023, §§ 61–63; Bharatiya Nyaya Sanhita, No. 45 of 2023, §§ 335–336; Information Technology Act, No. 21 of 2000, §§ 66C–66E, 67, 67A; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026.