

Deepfakes as Digital Evidence under Indian Law: Challenges of Authenticity, Admissibility, and AI-Based Manipulation

Maneesh Kundana^{1*} | Nidhi Kundu²

¹Assistant Professor, Law, Government Law College, Nawa, Didwana-Kuchaman, India.

²Assistant Professor, Computer Science, Sri Karan Narendra Agriculture University, Jobner, Jaipur, India.

*Corresponding Author: mkkundana88@gmail.com

Citation: Kundana, M. & Kundu, N. (2026). Deepfakes as Digital Evidence under Indian Law: Challenges of Authenticity, Admissibility, and AI-Based Manipulation. International Journal of Global Research Innovations & Technology, 04(03), 169–178.

ABSTRACT

The swift development of generative artificial intelligence technology has led to the production of realistic audio, images, and videos generated synthetically, which is known as deepfakes. Even though these technologies have some legitimate uses, their sophistication leads to serious issues regarding the validity of digital evidence presented during judicial proceedings. Deepfakes can manipulate facial expressions, voices, movements, and occurrences with such realism that the verification through visual inspection alone cannot establish its legitimacy. Thus, there arises an important evidentiary challenge where digital evidence that is realistic no longer presents a real person or occurrence. This paper discusses the evidentiary challenges raised by deepfakes to the legitimacy and admissibility of digital evidence within the context of Indian law with special focus on Bharatiya Sakshya Adhiniyam, 2023 and jurisprudence on electronic evidence. The paper also addresses the use of AI-based deepfake detection and digital forensics for the validation of the manipulated digital media. It is argued that, while legal admissibility criteria may suffice to mitigate AI-enabled manipulation to some degree, without additional technical verification, this approach is unlikely to be effective against complex forms of manipulation. Therefore, a techno-legal approach to authentication combining such factors as provenance verification, metadata and integrity checks, forensics, AI-supported deep fake detection, expert evaluation, and chain of custody protection has been suggested.

Keywords: Deepfake, Artificial Intelligence, Digital Evidence, Electronic Evidence, Bharatiya Sakshya Adhiniyam, Generative AI, Digital Forensics, Evidence Authentication.

Introduction

Digital evidence has become increasingly important in modern judicial proceedings as communication, financial transactions, surveillance, social interactions, and personal activities are routinely recorded through digital devices. Photographs, CCTV recordings, mobile phone videos, voice recordings, social-media content, emails, and other electronic records may provide important information for establishing facts in both civil and criminal proceedings. Traditionally, audio-visual material has often possessed strong persuasive value because it appears to provide a direct representation of a person, statement, action, or event. The emergence of artificial intelligence (AI)-generated synthetic media, however, increasingly challenges this assumption.

Deepfakes are digitally generated or manipulated media in which AI techniques are used to alter or synthesize visual or auditory content with a high degree of realism. Advances in deep learning have enabled the manipulation of facial identity and expressions, lip movements, speech, and even entire visual scenes (Mirsky & Lee, 2021; Tolosana et al., 2020). Deepfake generation has progressed from relatively obvious face-swapping demonstrations to sophisticated synthetic media that may be difficult for ordinary viewers to distinguish from authentic content. The increasing accessibility of generative AI tools further reduces the technical expertise and resources required to produce manipulated media.

The implications extend beyond misinformation and online deception. Deepfakes present a significant challenge to digital forensics and the administration of justice because manipulated media may potentially be presented as evidence of an event that never occurred or used to challenge genuine evidence by alleging that authentic material was artificially generated. Media-forensics research has therefore increasingly focused on identifying traces of manipulation through spatial, temporal, physiological, compression, and other digital artefacts (L. Verdoliva, 2020). Deepfake-detection systems based on machine learning and deep learning have also demonstrated considerable ability to identify manipulated facial media under controlled experimental conditions (Li et al., 2020; L. and R. C. and T. J. and R. A. and C. D. and N. M. Verdoliva, 2019). Nevertheless, their reliability may decrease when confronted with previously unseen manipulation techniques, compression, post-processing, low-quality recordings, or rapidly evolving generative models (Lyu, 2020).

The evidentiary challenge is therefore two-sided. On one hand, a convincing deepfake may be incorrectly accepted as authentic. On the other hand, the increasing public awareness of synthetic media may allow genuine recordings to be dismissed merely by claiming that they are AI-generated. Consequently, the existence of deepfake technology affects not only the detection of fabricated evidence but also the perceived reliability of authentic digital evidence.

In India, the legal treatment of electronic and digital evidence has undergone significant development. Earlier jurisprudence under the Indian Evidence Act, 1872, particularly *Anvar P.V. v. P.K. Basheer and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal* (*Anvar P.V. v. P.K. Basheer and Others*, 2014; *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal and Others*, 2020), emphasized statutory requirements governing electronic records and their admissibility. The Bharatiya Sakshya Adhiniyam, 2023 (BSA), which came into force on 1 July 2024, now expressly incorporates electronic and digital records within India's contemporary evidentiary framework. However, compliance with procedural requirements for admitting an electronic record does not necessarily establish that the content represented in an audio, image, or video has not been synthetically altered. Deepfakes, therefore, raise an essential distinction between the admissibility of digital evidence and the authenticity of the underlying information itself.

The distinction raises an important techno-legal challenge. For instance, while hash values and chain-of-custody processes can show that digital files have not been modified after their acquisition, they cannot independently establish the fact that the files were authentic when they were created or acquired. Equally, an AI-based detector can detect signs of manipulation but not determine the admissibility of evidence independently. It is only through a proper interaction between evidentiary law and digital forensics authentication that an effective evaluation can take place.

Against this backdrop, the paper explores the following research questions: (RQ1) What are the challenges in establishing the authenticity and reliability of digital evidence in the Indian context in light of deepfakes? (RQ2) How far is the existing Indian evidentiary law equipped to handle AI-generated and AI-altered audio-visual evidence? (RQ3) How can the process of AI-based forensic detection be integrated with legal provisions for strengthening authentication of suspected deepfake evidence?

The paper makes an original contribution in analyzing the problem of deepfakes with specific attention to Indian digital evidence and in developing a framework for techno-legal authentication. The suggested framework combines several elements: source and provenance checking, integrity checking, forensic analysis, AI-based detection, expert analysis, and safeguards for digital evidence, instead of suggesting that the use of AI for detecting deepfakes should be considered as a replacement of judicial or forensic evaluation of such evidence.

Deepfake Technology and Digital Evidence

Deepfake technology refers to the creation of artificial or manipulated digital media with the help of advanced artificial intelligence tools that produce real-life digital representations of people, their voices, expressions, and actions. Earlier technologies utilized autoencoders and GANs for deepfake creation, while newer generations of generative technologies have provided more opportunities for high-quality and easy-to-use synthetic materials (Mirsky & Lee, 2021; Tolosana et al., 2020). The process of manipulation can include face swapping, facial reenactment, lip sync, voice duplication, or generation of the entire piece of synthetic media. In terms of evidence, deepfakes are crucial because they may alter identity, speech, behavior, or even existence of a particular individual in the digital media.

• **Forms of Deepfake Manipulation**

Face swapping technologies swap the face of one individual in a pre-existing image or video with that of another while trying to maintain realistic facial expressions, pose, and lighting. Facial reenactment edits expressions or motions without necessarily swapping the identity of the individual. The alteration of lip-sync changes the mouth motion to match a different audio, thus making the individual appear as if he or she stated something which he or she did not actually say. Likewise, artificial intelligence-based voice cloning technology produces speech with voice characteristics matching those of the target individual. More complex generative technologies can create synthesized images, audio, and videos without authentic source materials.

All the aforementioned techniques have implications for digital evidence. An edited CCTV footage may make it look like an individual was at a certain place, an edited video may show wrong actions by an individual, and an edited audio may show an authentic looking conversation or statement by an individual. As such, the issue regarding digital evidence is not just whether the digital file has been edited. It is also whether the content represented in the file matches a real-life event.

Table 1: Major forms of AI-based manipulation and their evidentiary implications

Manipulation Type	Description	Potential Evidentiary Risk	Possible Forensic Indicators
Face swapping	Replaces one person's face with another	False identification or presence	Facial inconsistencies, blending artefacts, temporal anomalies
Facial reenactment	Alters expressions or facial movements	Misrepresentation of behaviour or reactions	Motion and landmark inconsistencies
Lip-sync manipulation	Modifies lip movements to match different speech	Fabricated statements	Audio-visual synchronization anomalies
Voice cloning	Generates speech resembling a targeted speaker	False confession, conversation, or instruction	Spectral, acoustic, and prosodic anomalies
Fully synthetic media	Generates artificial images, audio, or video	Fabrication of an event that never occurred	Generative artefacts, provenance inconsistencies, forensic model outputs

• **Deepfake Detection and Digital Forensics**

Digital forensic methods strive to establish the manipulative presence in media by analysing features that would not be easily visible to a human eye. Media forensics involves the analysis of metadata, compression, sensor properties, editing properties, and inconsistencies that occur in the manipulation process. In connection with the use of deepfakes, many artificial intelligence-driven solutions have been proposed for detecting subtle patterns and inconsistencies of synthetic media (Verdoliva, 2020).

Detectors based on the deep learning method may study facial texture, blending boundaries, unusual illumination, physiological signals, frequency-domain properties, or inconsistencies across video frames (Qureshi et al., 2024). Such benchmark datasets as FaceForensics++ and Celeb-DF have been instrumental in creating and testing the effectiveness of those systems (Li et al., 2020; L., R. C., T. J., R. A., C. D., and N. M. Verdoliva, 2019). The datasets prove that the machine-learning model can be trained to detect discriminative features of media manipulation and can provide high performance on specific datasets (Gong & Li, 2024).

However, merely considering the effectiveness of the AI detection system as evidence of the authenticity of content may be misleading. The issue of generalization often appears with deepfake detection algorithms: a model, which is based on certain manipulation techniques or data sets, might work less efficiently when dealing with different kinds of manipulations that were not present in training data sets (Khan & Dang-Nguyen, 2023). Compression, resizing, processing through social media, noise, low-quality video surveillance footage, and efforts to avoid detection are also possible. Furthermore, as generative models evolve, the artefacts of both types may become less noticeable.

This leads to the technological race between generative and forensic analysis. Advancements in the former drive improvements in the latter and vice versa. In turn, depending only on one AI detector as a proof of authenticity of the evidence raises some reliability issues.

- **From File Integrity to Content Authenticity**

It is, thus, necessary to make a key distinction between file integrity and content authenticity. Integrity mechanisms, such as cryptography hashes, allow determining whether the given file has been changed from some point of acquisition onwards. Proper chain-of-custody documentation can also reveal information regarding the process of acquisition, handling, storage, and examination of digital evidence. These mechanisms are crucial for ensuring that evidence remains protected from any further alterations.

Nonetheless, an unaltered file does not necessarily mean that it is an authentic representation of reality. If some AI-generated or manipulated video is acquired and its hash is calculated at once, the given hash will prove that no changes have occurred later, but it will not prove that the given event took place. Namely,

Integrity asks: "Has this digital file changed since acquisition?"

Authenticity asks: "Does this digital content truly represent the person, statement, action, or event that it claims to represent?"

The use of deepfakes only renders this difference relevant to judicial consideration. Authentication of potentially fake media might involve different methods which include source or provenance evaluation, meta data examination, integrity checks, forensic analysis, artificial intelligence tools for manipulation detection, and interpretation. All technical results should then be examined in the context of the appropriate rules of evidence and not taken as a separate assessment of admissibility.

In this way, the main problem caused by deepfakes does not involve mere detection of manipulated pixels or audio files. The problem relates to establishing sufficient reliability of the link between a digital recording and the actual occurrence depicted by it. Such an issue is especially relevant in cases when this type of material is presented as evidence in court.

Indian Legal Framework Governing Digital Evidence

There have been changes in the law regarding electronic evidence in India due to the increasing use of digital documents in legal cases. In particular, the Bharatiya Sakshya Adhiniyam, 2023 (BSA) (The Bharatiya Sakshya Adhiniyam, 2023, 2023), which replaced the Indian Evidence Act of 1872, constitutes the present-day legislative basis for electronic and digital evidence. Under the BSA, both electronic and digital records have been explicitly acknowledged and incorporated into the general scheme of documentary evidence. The acknowledgment is especially important considering deepfakes since the manipulated photographs, audio files, video files, CCTV footage, and other digital evidence could still be used as electronic evidence.

- **Electronic and Digital Records under the BSA**

Section 61 of the BSA provides that an electronic or digital record cannot be denied admissibility merely because of its electronic or digital nature. Subject to Section 63, such a record has the same legal effect, validity, and enforceability as other documents (Chesney & Citron, 2019). Section 62 further provides that the contents of electronic records may be proved in accordance with Section 63. Section 63 consequently assumes particular importance in determining the admissibility of electronic records produced from computers and other electronic systems.

Section 63 broadly addresses circumstances in which information contained in an electronic record, including information printed, stored, recorded, or copied in electronic form, may be treated as a document when the statutory conditions are satisfied. These conditions are directed toward establishing the reliability of the computer or communication system and the manner in which the electronic record was produced. The provision also requires a certificate containing prescribed particulars when an electronic record is submitted for admission in accordance with the section (Chesney & Citron, 2019).

The BSA therefore provides an important statutory mechanism for establishing the evidentiary basis of electronic records. However, deepfake technology introduces a separate question that may arise even where the procedural requirements governing an electronic record have been satisfied: whether the content contained in the record is itself genuine.

For instance, let us take the example of the video that is received in a digital device and is later on stored as per the proper procedure. Fulfillment of the conditions in respect of the electronic document might help in proving the source or creation of the copy produced in front of the court. It is not necessary that it proves that the speech, face, or incident in the video were not fabricated before the creation of the file in the evidentiary trail.

- **Judicial Development of Electronic Evidence**

Prior to the implementation of the BSA, there were numerous judicial deliberations on electronic evidence under Sections 65A and 65B of the Indian Evidence Act of 1872 by the Supreme Court of India. There were two significant judgements regarding the development of the principles applicable to electronic records.

Firstly, in *Anvar P.V. v. P.K. Basheer & Ors.* (2014), the Supreme Court stressed the statutory procedure for the use of electronic records and ruled that the electronic evidence in the form of secondary electronic records should meet the criteria stipulated in Section 65B of the Indian Evidence Act [12]. This judgement increased procedural protection applicable to electronic evidence since statutory compliance was stressed in the event that electronic records were used in judicial procedures.

Later on, the issue of electronic records and their admissibility was considered in detail in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal & Ors.* (2020). In this case, the Supreme Court reemphasized the principles established in *Anvar P.V.* and stressed the significance of the certificate stipulated in Section 65B(4) concerning electronic records that were produced by secondary means (The Information Technology Act, 2000, 2000).

These rulings were issued according to the Indian Evidence Act, 1872, but still have relevance when it comes to the development of doctrine related to the authentication of electronic evidence in India. Currently, the BSA is the statute that governs this process, and sections 61–63 of this statute continue to demonstrate the significance of reliability.

- **The Deepfake Authentication Gap**

Deepfakes may pose a possible mismatch between the legal processes of establishing admissibility of an electronic record and the technical processes for proving its authenticity. There are three distinct issues that may be considered where manipulation by AI is thought to have taken place:

- Source: From where was the electronic record obtained?
- Integrity: Has the electronic record been altered since it was obtained or stored?
- Authenticity: Is the content truly representing the person, speaking, act, or event that it claims to represent?

Issues (i) and (ii) are possible to demonstrate using evidence about the device, method of obtaining it, chain of custody, hash values, system logs, and statutory certification. Issue (iii), however, is far more problematic in the case of sophisticated synthetic media.

For example, an artificial video can be created by means of an AI technology and then saved in the computer system. Then, a copy of the video may be created in accordance with the relevant requirements related to the electronic record. A hash may also confirm that the examined copy of the video is the same file as was initially obtained. Neither of these two facts proves that the event presented in the video took place. The process of authentication of the evidence, thus, should involve separation of the authenticity of the electronic record as a file from the authenticity of the content.

Moreover, such distinction is needed in relation to the assessment of expert evidence. According to s. 39 of the BSA, the expert opinion is a valid source of proof where special expertise is required. Therefore, in a dispute involving deepfake, digital forensics experts may be of significant importance when conducting examination of the metadata, coding features, signs of manipulation, inconsistencies in audio and visual data, device data, and results of analysis performed by deepfake detector. However, an expert opinion cannot be based only on the probability score produced by the system.

- **Need for a Layered Evidentiary Approach**

Thus, the problem of the legality of the deepfakes issue does not merely concern itself with the admissibility of the electronic evidence. The courts might find themselves in the situation where the party opposite would be contesting the admissibility of the evidence based on the grounds that the evidence was manipulated or that it is not authentic but rather a deepfake.

Hence, an appropriate response should take care of avoiding both extremes of accepting any digital evidence merely because of its credibility and rejecting the same on the basis of mere existence of the technology. If there is a reasonable doubt regarding manipulation, the authentication process must rely on various sources of verification, including provenance, integrity, circumstantial evidence, digital

device and metadata examination, forensic analysis, expert witness testimony, and artificial intelligence-assisted deepfake identification techniques.

Consequently, the BSA lays down the necessary legislation regarding the electronic and digital evidence, while the development of synthetic media technology poses an extra forensic issue of the veracity of the information in question. Solving this issue involves the evidentiary framework functioning together with the technologically sound mechanisms of the digital authentication process.

Challenges of Authenticity and Admissibility of Deepfake Evidence

In relation to the increasingly advanced use of synthetic media, the evidentiary issues go beyond the traditional evidentiary problems related to the alteration of electronic evidence. With the deepfake, the problem may not lie in whether a file has been altered or not since one can raise questions as to whether the recorded person, speech, action or event has ever occurred in the form reflected in the media file. The process poses many problems for verification, authentication and attribution purposes.

- **Establishing the Authenticity and Provenance**

Authentication involves the reliable linkage of digital evidence with the events or individuals it represents. Deepfakes present a challenge in this regard since realistic audio-visual recordings can be made without any corresponding action being taken in the real world. Thus, visual realism can no longer be a reliable criterion of authentication.

Provenance information is very useful in this situation. Information on the originating source, timing of generation, manner of collection, metadata, post-collection manipulation, and the transfer history of the digital information can help in the determination of the reliability of digital evidence.

- **Chain of Custody and Pre-existing Manipulation**

The appropriate chain of custody in relation to digital evidence involves documentation of possession, transfer, storage, and analysis since acquisition. Hash values using cryptography could also provide evidence if the file collected had undergone any modifications thereafter.

However, deepfakes illustrate another deficiency of such measures for preservation. Chain of custody starts at the stage of collection. In case the evidence was already manipulated at the time of collection, an impeccable chain of custody would only show that the file manipulated was preserved.

For instance, an artificial audio file might be generated through voice cloning and then seized from a digital device. Investigators might generate a hash value and secure it. As a result, a chain of custody proves that the file did not change but cannot prove that the conversation ever took place. Thus, preservation and authenticity have to stay separate concepts.

- **Reliability and Explainability of AI-Based Detection**

The use of AI-powered detectors to identify fake videos can be considered an essential mechanism; however, using their outputs as evidence causes certain concerns. Detection algorithms are generally trained on certain data sets, incorporating certain manipulation methods. They tend to perform poorly when exposed to new generators, unknown data sets, compression, low resolution, or even adversarial changes.

One more issue connected with the use of AI detection technology is explainability. An algorithm which identifies a certain video as "deepfake" with certain probability provides valuable forensic information. However, just a number does not reveal what makes this video deepfake and if these features are enough to consider this video as deepfake in this particular case. In the context of the court proceeding, the underlying methodology may have to be proved.

Thus, AI-powered detectors have to be used as an investigatory and forensic assistance tool, and not as an independent evidence source.

- **Attribution of AI-Based Manipulation**

Even in cases where forensics proves the manipulation of the media, another problem arises. The mere discovery that a video has been made through deepfaking technology does not imply that one knows who created the video, who spread it, or whether a specific individual has tried to deceive anyone.

This issue could be relevant in cases where culpability is dependent on knowledge, intention, authorship, or distribution. In other words, apart from establishing whether or not a deepfake was involved, it may take more evidence to determine whether someone is guilty of producing and using the media.

- **The “Liar’s Dividend” and Genuine Evidence**

One of the most challenging problems resulting from the creation of deepfakes is that their very existence may put in question not only deepfaked evidence but real evidence too. Once deepfakes become common knowledge, a person who is presented with some real evidence will always be able to claim that it was made with the help of artificial intelligence.

Such a strategy is called the liar’s dividend – the possibility to use the presence of convincing deepfakes as a tool for calling into question any recordings. In this way, the use of an evidentiary approach that treats every allegation of the manipulations of deepfakes as enough to disqualify digital evidence may undermine the administration of justice itself.

Courts should thus manage two possible mistakes: accepting the manipulated evidence as real and rejecting the real evidence as fake. Mere claim about the artificial origin of some recordings shouldn’t be enough to decide about their evidentiary weight. However, at the same time, credible challenges to it may have certain grounds for additional analysis.

- **Evidentiary Weight in the Age of Synthetic Media**

Ultimately, the issue of deepfakes highlights the reason why the concepts of admissibility and evidentiary weight are different in the first place. The electronic record might comply with all the procedure-related criteria necessary for being admitted into evidence in the court proceedings; at the same time, concerns related to authenticity, reliability, or interpretation of such record might affect the weight of the latter.

Consequently, the process of assessing deepfake evidence should be context-specific rather than a matter of black-or-white. In particular, some factors that should be considered are source of the recording, chain of custody, metadata, information about the device, testimony, consistency with other evidence, forensics results, and reliability of the AI technology itself.

Therefore, the question is not to assume that all digital recordings are unreliable, but to devise a mechanism of authentication that would be able to distinguish real digital evidence from the sophisticatedly created fake one.

Proposed Techno-Legal Framework for Authentication of Suspected Deepfake Evidence

As can be seen from the above analysis, the legality and use of technological means of deepfake detection independently cannot be the proof of the validity of suspected synthetic media. In view of the above, this paper introduces a framework for Techno-Legal Authentication of Deepfake Evidence called TL-DEAF. It involves measures of authentication of evidence together with digital forensics and AI-based verification of the evidence as illustrated in Figure 1.

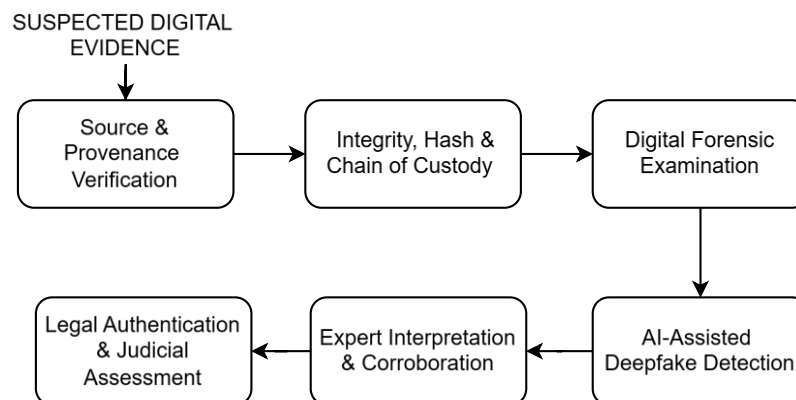


Figure 1: Proposed Techno-Legal Deepfake Evidence Authentication Framework (TL-DEAF)

- **Layered Authentication Framework**

Stage 1: Source and Provenance Verification

The first stage establishes the origin and history of the digital material. Investigators should identify, wherever possible, the originating device, creator, date and time of creation, acquisition method, transmission history, and available provenance information. Emerging content-provenance mechanisms may further assist in recording the origin and modification history of digital media.

Stage 2: Integrity and Chain-of-Custody Verification

After acquisition, the digital evidence should be preserved in its original form and cryptographic hash values should be generated. Every subsequent transfer, access, duplication, or forensic examination should be documented. This stage establishes that the evidence presented for examination corresponds to the material originally acquired and has not been altered during evidentiary handling.

Stage 3: Digital Forensic Examination

The evidence should then undergo forensic analysis of metadata, encoding characteristics, compression patterns, file structure, temporal consistency, audio-visual synchronization, and other relevant indicators. Examination at this stage can identify anomalies that may indicate editing, synthesis, or manipulation.

Stage 4: AI-Assisted Deepfake Detection

Where synthetic manipulation is reasonably suspected, one or more validated AI-based detection techniques may be applied. Depending on the type of evidence, these systems may analyse facial artefacts, spatial and temporal inconsistencies, frequency-domain characteristics, acoustic patterns, or other manipulation indicators. Because detector performance may vary across datasets and previously unseen generation techniques, an AI-generated probability or classification should not independently determine authenticity.

Stage 5: Expert Interpretation and Corroboration

Forensic findings must be assessed by properly qualified forensic experts. The expert must take into account the method used for detection, the limitations of the model used, the quality of the analyzed evidence, the risk of false positives and negatives, and the consistency of various forensic indicators. Forensic findings should also be supported, when feasible, with other independent sources of information, including witness accounts, data from devices, location data, communication data, or other contemporary materials.

Stage 6: Legal Authentication and Judicial Assessment

The last stage pertains to the relevant evidentiary rules contained in the Bharatiya Sakshya Adhinyam, 2023. The court must consider statutory admissibility along with the provenance, integrity, forensic findings, expert assessment, and corroborating circumstances. Accordingly, the AI-generated output should provide assistance, but not substitute the judicial determination.

- **Suggested Workflow**

The suggested framework may be described as follows:

One significant benefit of the framework is that failure or uncertainty at one level of analysis does not automatically determine the result. For example, lack of credible provenance of evidence does not automatically mean its fabrication, in the same way that a positive test result of a deepfake detector does not automatically confirm manipulation.

- **Policy and Procedural Recommendations**

The implementation of the proposed framework will require increased collaboration among the legal process, digital forensics, and governance of artificial intelligence. The Indian investigative and forensic agencies should establish standard practices for dealing with suspected evidence created by artificial intelligence technology, such as preservation of the original files, cryptographic hash computation, provenance documentation, and replicable forensic process. In cases where artificial intelligence-based detectors were employed, the forensic report must indicate the name of the tool or model, the purpose for its use, its limitations, and grounds for interpretation of the output of the detector.

Moreover, courts must differentiate between machine-generated results and expert analysis. The result generated by artificial intelligence is not conclusive evidence simply because it was obtained by a highly sophisticated tool. Human experts have to interpret the technical findings and explain their evidentiary value.

Finally, provenance technology and authenticated creation tools should be encouraged where possible. They could help post hoc detection of deepfakes by providing information about the source and editing process of digital media. However, lack of provenance credentials should not automatically discredit the evidence, especially if it concerns legacy content and media that cannot use those technologies.

The suggested TL-DEAF system thus follows the layered verification paradigm, rather than a single tool based detection approach. Through the integration of provenance, integrity verification, forensic analysis, AI-based detection, human expert evaluation, corroborative evidence, and legal analysis, a more balanced approach to deepfake evidence is ensured while maintaining the evidentiary value of digital evidence.

Discussion and Legal–Policy Implications

It has thus been established that the problem of fake evidence is a mixture of technological and legal issues. Whereas the Bharatiya Sakshya Adhiniyam, 2023 offers statutory guidance in the use of electronic and digital records, the advancement of the technology behind AI-created media calls for authentication not just of the creation and storage of an electronic record, but its content – the fact whether the content of the record represents a real person, statement, action or occurrence.

In this regard, the offered TL-DEAF framework relies on a layered approach instead of a singular way of authentication of such records. It is especially crucial due to the presence of certain weaknesses in the use of any possible method of authentication. Thus, for example, while cryptographic hash functions allow proving file integrity after acquisition of data, they do not prove the authenticity of the initial content of such records. Metadata might be used to prove provenance, although incomplete and changed. In addition, AI-based deepfake detectors may prove the presence of manipulations, but may show different results if facing an unknown generator or low quality audiovisual content.

Content-provenance technologies provide an additional opportunity for strengthening digital authentication. Standards such as the Coalition for Content Provenance and Authenticity (C2PA) can cryptographically associate provenance information with digital content, allowing information concerning its origin and modification history to be examined. However, provenance should be understood as complementary to, rather than a replacement for, digital forensic analysis. A valid provenance record can demonstrate that particular assertions associated with an asset have remained intact, but it does not independently establish the truth of the real-world event depicted in the content.

Three practical implications follow from this analysis. First, forensic laboratories and investigating agencies should develop standardized procedures specifically for suspected AI-generated or AI-manipulated evidence. Such procedures should emphasize preservation of original media, cryptographic hashing, provenance examination, documentation of the chain of custody, and reproducibility of forensic findings.

Second, where AI-based deepfake detectors are relied upon, their outputs should be accompanied by information concerning the detection method, version or model used, testing conditions, limitations, and confidence associated with the conclusion. Courts should avoid treating a machine-generated classification such as “real” or “fake” as conclusive without appropriate expert interpretation.

Third, judicial and forensic capacity-building will become increasingly important. Judges, investigators, lawyers, and forensic experts need sufficient understanding of synthetic-media technologies to distinguish questions of admissibility, integrity, authenticity, and evidentiary weight. At the same time, safeguards should prevent unsupported allegations of AI manipulation from automatically discrediting genuine recordings.

The appropriate policy response is therefore neither automatic trust nor automatic distrust of digital evidence. Instead, the evidentiary system should adopt a proportionate and technically informed approach in which the level of forensic scrutiny reflects the nature of the authenticity challenge and the significance of the evidence to the proceeding.

Conclusion

Deepfake technology fundamentally challenges the traditional assumption that realistic audio-visual material necessarily corresponds to a real event. As generative AI becomes increasingly capable of synthesizing faces, voices, speech, and actions, establishing the reliability of digital evidence requires more than determining whether a file has remained unchanged after acquisition.

This paper examined the challenges posed by deepfakes to authenticity and admissibility within the Indian evidentiary framework, particularly under the Bharatiya Sakshya Adhiniyam, 2023. The analysis revealed a vital distinction between admissibility, integrity, and content authenticity. While compliance with legislation allows an electronic record to be treated as admissible, it may require further forensic analysis to ascertain whether its content was synthetically created or altered.

To fill this lacuna, the paper developed the Techno-Legal Deepfake Evidence Authentication Framework (TL-DEAF). The framework involves the following steps: source and provenance verification; integrity and chain-of-custody evaluation; digital forensics; artificial intelligence-based deepfake detection; expert evaluation and corroboration; and finally, legal assessment. It is deliberate to consider AI-based detection as complementary to forensic expertise and judicial assessment.

The future reliability of digital evidence is likely to depend on cooperation between law and technology. Standards of provenance, improved forensic tools, procedures, capacity building and judicial understanding can all contribute to the authentication process. Nevertheless, technological tools must serve as a supplementary tool, and not the decisive factor. A combination of legal and technological layers is a more balanced way of protecting courts from sophisticated synthetic evidence without devaluing digital evidence simply because convincing deepfakes are now technically possible.

References

1. Anvar P.V. v. P.K. Basheer and Others, 10 Supreme Court of India 473 (2014).
2. Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal and Others, 7 Supreme Court of India 1 (2020).
3. Chesney, R., & Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107(6), 1753–1820.
4. Gong, L. Y., & Li, X. J. (2024). A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges. *Electronics*, 13(3), 585. <https://doi.org/10.3390/electronics13030585>
5. Khan, S. A., & Dang-Nguyen, D.-T. (2023). *Deepfake Detection: A Comparative Analysis*. <https://doi.org/10.48550/arXiv.2308.03471>
6. Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3204–3213. <https://doi.org/10.1109/CVPR42600.2020.00327>
7. Lyu, S. (2020). Deepfake Detection: Current Challenges and Next Steps. *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 1–6. <https://doi.org/10.1109/ICMEW46912.2020.9105991>
8. Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys*, 54(1), 1–41. <https://doi.org/10.1145/3425780>
9. Qureshi, S. M., Saeed, A., Almotiri, S. H., Ahmad, F., & Al Ghamdi, M. A. (2024). Deepfake Forensics: A Survey of Digital Forensic Methods for Multimodal Deepfake Identification on Social Media. *PeerJ Computer Science*, 10, e2037. <https://doi.org/10.7717/peerj-cs.2037>
10. The Bharatiya Sakshya Adhiniyam, 2023 (2023).
11. The Information Technology Act, 2000 (2000).
12. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. *Information Fusion*, 64, 131–148. <https://doi.org/10.1016/j.inffus.2020.06.014>
13. Verdoliva, L. (2020). Media Forensics and DeepFakes: An Overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910–932. <https://doi.org/10.1109/JSTSP.2020.3002101>
14. Verdoliva, L. and R. C. and T. J. and R. A. and C. D. and N. M. (2019). Faceforensics++: Learning to detect manipulated facial images. *Inte. Conf. on Comp. Visi.(ICCV). IEEE/CVF*.

