ADMISSIBILITY AND RELIABILITY OF ARTIFICIAL INTELLIGENCE-GENERATED EVIDENCE IN INDIAN COURTS

Author- Nikita Desai,
Jaipur National University

ABSTRACT

Artificial intelligence is not only changing how evidence is collected and analyzed, but also increasingly how such evidence is being created. Deepfake videos, generated sounds, AI-created visuals, predictive algorithms, results from facial recognition, automated forensics reports and other types of generated AI content pose new challenges for Indian courts that go beyond the traditional challenges posed by electronic evidence. The Bharatiya Sakshya Adhiniyam, 2023 ('BSA') recognizes electronic and digital evidence and provides for the conditions under which such evidence is to be admitted in court. However, the mere adherence to the procedures necessary for showing that a certain electronic document was created on a computer does not mean that the document is true and reliable.

This paper considers the question of whether or not AI-generated evidence can be admitted and relied upon as part of the current Indian evidentiary framework. It seeks to analyze the shift from the Indian Evidence Act, 1872 to the BSA and questions the relevance of the Supreme Court's electronic evidence jurisprudence, especially with regard to Anvar P.V. v. P.K. Basheer and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal. The significance of the distinction between admissibility and weight of evidence is highly relevant when dealing with evidence generated by AI technologies.

The paper further examines the constitutional issues arising out of the opacity of algorithmic evidence in Article 14 and Article 21, where the accused is unable to probe the methodology behind the incriminating evidence presented. With the aid of comparative law, the paper advances an Artificial Intelligence Evidence Reliability Test ("AIERT") for use in the Indian courts. This framework assesses provenance, authenticity and integrity, extent of AI involvement, system reliability and error rate, explainability, reproducibility, human oversight, independent corroboration, and the opportunity for adversarial challenge. The paper finds that India should not establish a presumption against AI evidence; on the contrary, it should have a neutral approach to evidence technology, alongside a rigorous reliability test for AI involvement.

Keywords: Artificial Intelligence; Electronic Evidence; Bharatiya Sakshya Adhiniyam; Deepfakes; Algorithmic Evidence; Admissibility; Reliability; Fair Trial.

I. INTRODUCTION

The law of evidence operates on what appears to be an elementary premise: evidence stands for, records or transmits something which existed independently of the process of its delivery. A photograph is taken of a scene that exists; sound is recorded that has been spoken; data is stored on a computer that has already been created or inputted. This premise is challenged by artificial intelligence. Modern forms of AI technology can produce a picture of an event that never took place, recreate a voice saying words that were never said, fill gaps in video footage, recognize an individual based on probabilistic facial recognition, or draw conclusions from datasets using processes that even the AI's operators cannot fully understand.

This issue is thus no longer limited to determining whether there has been any alteration to the electronic document after its creation but whether the content itself is an authentic creation. There are great implications of such distinction under Indian law of evidence. The Bharatiya Sakshya Adhiniyam, 2023, which is the successor to the Indian Evidence Act, 1872 ("IEA"), recognizes electronic and digital records explicitly. Sections 61 to 63 form the main provisions relating to electronic evidence. Though the BSA is more technology oriented than the nineteenth century statute, its electronic evidence system has mostly remained similar to the one developed even before the advent of generative AI and machine learning systems.[1]

Similarly, the important Supreme Court judgments in relation to electronic evidence have arisen mostly from cases involving recording devices, CDs, telephone call records, CCTVs and computer output. For example, in *Anvar P.V. v. P.K. Basheer*, it was emphasized that the provisions for electronic records have to be adhered to as per the special law enacted in this regard. Likewise, in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*, the mandatory nature of the requirement for certification of secondary electronic evidence was reiterated.[2][3]

In any case, however, AI creates a secondary problem. Let's say a deepfake video is created using an AI program, which is then perfectly preserved on the computer. The hash value of the video does not change; its chain of custody is perfect; all of the conditions for creation of the electronic

[1]Bharatiya Sakshya Adhiniyam, No. 47 of 2023, §§ 61-63, INDIA CODE (2023).
[2]Anvar P.V. v. P.K. Basheer, (2014) 10 S.C.C. 473 (India).
[3]Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 S.C.C. 1 (India).

file are met. This ensures that the file presented to the court is identical to the one which has been collected. However, it does not guarantee that anything shown in this file really happened.

The present essay, accordingly, makes two propositions that are related to each other. First, there must be a distinction between electronic authenticity and substantive reliability while dealing with evidence produced by AI in courtrooms. Secondly, an India-centric test for evaluating reliability of AI evidence is needed in addition to existing requirements in BSA in respect of electronic evidence. Towards this end, the present essay proposes the Artificial Intelligence Evidence Reliability Test (AIERT), which is made up of nine factors.

II. UNDERSTANDING "AI-GENERATED EVIDENCE"

"AI-generated evidence" must not be considered as an umbrella term. Rather, three separate kinds need to be distinguished. Firstly, there is evidence generated by AI. The AI may generate synthetic text, sound recordings, photographs and videos. Deepfakes serve as an example. The main issue with such evidence is its accuracy with respect to the real world. Secondly, there is AI-aided evidence. In this case, the real-life evidence exists, however, it undergoes a transformation, reconstruction or enhancement due to the use of AI. For instance, this may refer to the improvement of the resolution of CCTV video, reduction of background noise in audio, filling the gaps in videos, enhancing the facial features using generative AI and others. At the same time, such methods increase the evidential value while adding new information to the source material. Thirdly, there is AI or algorithmic evidence. AI here may or may not generate the evidence but makes an inference from the data. This category includes facial recognition, predictions, handwriting comparison, etc. This is important, for reasons discussed below, because the issue of reliability varies for each of the scenarios. When it comes to deepfakes, the key issue becomes that of authenticity. When it comes to AI enhancement, the court needs to find out what parts of the information were recorded and what parts were inferred through the use of algorithms. In case of classification by algorithm, attention will need to be paid to the methodology, accuracy, data, error rates, and reproducibility.

III. THE INDIAN STATUTORY FRAMEWORK

A. Electronic Evidence under the Bharatiya Sakshya Adhiniyam

The BSA, which came into force on 1 July 2024, is the most important and relevant legal regime for evidence in India. According to Section 61, an electronic record cannot be made inadmissible

only because it is in electronic form. Section 62 says that the contents of electronic records may be proved in accordance with Section 63.[4][5]

The provisions contained in Section 63 set out the conditions for admissibility of computer-generated output. The section substantially carries forward the legislative scheme of the erstwhile Section 65B of the IEA, namely that electronic information may be treated as a document upon satisfaction of certain statutory conditions.[6]

Such safeguards still matter. Electronic evidence can be duplicated, modified, and sent out without leaving behind the physical evidence normally left by tampering with traditional documents. Technical and certification details might help courts trace back the origin and the integrity of an electronic record. However, the advent of AI points to the conceptual weakness of such safeguards. The questions raised by Section 63 are mainly focused on the computer and the creation of an electronic record. In case of the AI, the court will need to consider additional inquiries concerning the process of creating informational contents reflected in the electronic record. While a hash can prove that a certain file has not been modified after its retrieval, it cannot prove that the contents of the file were not artificially created in the first place. Hence, integrity does not equal truth.

B. Anvar P.V. and the Authentication Imperative

It would not be an exaggeration to say that the judgment by the Supreme Court in the case of *Anvar P.V. v. P.K. Basheer* was a turning point in the area of electronic evidence law in India. The Court ruled that the special laws dealing with electronic evidence would govern the admissibility of secondary electronic evidence, and general laws regarding secondary documentary evidence could not be substituted for the special provisions. However, the reasoning behind the judgment is still relevant in the AI era. Evidence in digital form is particularly prone to manipulation. Thus, procedural protection is necessary to shield judicial proceedings from records which cannot be proved to have originated from the source. Nevertheless, in Anvar, the electronic evidence was relatively conventional in nature. The current problem lies beyond.[7]

[4]Bharatiya Sakshya Adhiniyam, No. 47 of 2023, § 62, INDIA CODE (2023).
[5]Bharatiya Sakshya Adhiniyam, No. 47 of 2023, § 61, INDIA CODE (2023).
[6]Id. § 63.
[7]Anvar P.V. v. P.K. Basheer, (2014) 10 S.C.C. 473, 484-86 (India).

C. Arjun Panditrao and Procedural Reliability

In *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*, a judgment of a three-Judge Bench, put to rest the doubts that had arisen about the need for certification of secondary electronic evidence, overruling the contrary view taken in Shafhi Mohammad v. State of Himachal Pradesh and restoring the position in Anvar P.V. Although the ruling increased procedural clarity, it highlights another crucial distinction that is at the core of this essay. The certification process clarifies doubts about how the electronic document was created and verified for the purposes of being admissible in court. The process does not make everything stated in the certified document true or reliable. Anvar and Arjun Panditrao must therefore be treated as the first step, and not the last, in validating AI-generated evidence.[8]

IV. THE ADMISSIBILITY-RELIABILITY DIVIDE

Admissibility and reliability, though they overlap to a degree, remain distinct inquiries. In fact, Anvar recognised that truthfulness and reliability are issues quite separate from the issue of admissibility. This issue takes on great significance when one considers artificial intelligence.[9]

Imagine an AI-produced audio recording depicting an accused person making an admission about committing a crime. Law enforcement officials obtain the recording electronically and secure it through proper forensics practices. The prosecution will be able to determine the source of the recording, its receipt date, storage location, hash value, and lack of alterations made after its acquisition. This proves the integrity of the file. It does not prove the authenticity of the confession.

This problem occurs in a more subtle manner when using AI for enhancement. An algorithm might attempt to enhance a poorly-defined face on the video footage taken by a CCTV camera. Enhancement techniques in general could make the available information clearer. However, generative enhancement could generate missing features based on what is expected to be there. The output of the algorithm can even be convincing since the computer algorithm has generated data that never existed before.Courts must thus differentiate between source genuineness, document integrity, process integrity, and propositional integrity. Conventional methods of authenticating electronic evidence deal mainly with the first two. The new form of evidence requires consideration of all four.

[8]Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 S.C.C. 1 (India).
[9]Anvar P.V. v. P.K. Basheer, (2014) 10 S.C.C. 473 (India).

V. DEEPFAKES AND THE COLLAPSE OF VISUAL AUTHENTICITY

Audiovisual evidence has an odd potency due to the human predisposition to connect seeing with believing. The generative AI takes full advantage of this predisposition. Deepfake uses machine learning to generate or modify audiovisual evidence such that people seem to do things and say things that have not actually happened. What is dangerous about deepfakes goes far beyond the inclusion of the fabricated evidence. The mere knowledge that deepfakes can be convincing poses another problem: the real evidence might be disbelieved as being artificial.

This places Indian courts in a double bind. Either the fabricated evidence might be accepted as being genuine, or the real evidence might be rejected simply because fabrication was technologically possible. Neither situation provides an excuse to categorically exclude all digital audiovisual evidence. Rather, the courts need to insist more on provenance evidence. Provenance evidence would consist of the origin device, the original metadata, the acquisition chain, the cryptographic hash, the preservation methods and the availability of the original audiovisual material. However, importantly, the court should not see AI deepfake detection technology as an ultimate fix for the problem of deepfakes. The process of detection itself is based on the use of machine learning algorithms that might fail when confronted with new generation methods.

VI. ALGORITHMIC OPACITY, BIAS AND ERROR

A. The Black-Box Problem

Some advanced machine learning algorithms are hard to interpret in human terms. An algorithm may give a correct output without explaining the rationale behind the output based on certain attributes. This may be accepted in commercial scenarios, but in criminal proceedings, it is a completely different scenario. In case the evidence provided by an algorithm plays any significant role in depriving someone of their freedom, then that individual should usually have an opportunity to challenge the same.

B. Training-Data Bias

Pattern recognition in artificial intelligence occurs through learning from the data set. In cases where the data sets used to train AI inadequately represent certain populations, there will be disparities in their use. The accuracy of the entire process may be high enough, but the error rates in each group will vary greatly. The significance of this in terms of the law is clear. The court

would not be able to make any meaningful determination on the evidential value of an identification by an algorithm on the basis of the vendor's claim that it is "99% accurate."

C. Hallucinations and Generative Systems

The output from generative AI can be plausible yet untruthful information. The issue is one that presents a clear threat when summaries, transcriptions, translations, or reproductions done by AI are portrayed as an exact mechanical replication of the original source. In any situation where generative AI is used to convert the original source into evidence, the source itself should always be available for reference.

VII. CONSTITUTIONAL DIMENSIONS: ARTICLES 14 AND 21

The issue of evidence in terms of AI is not simply a question of the law of evidence per se. In any criminal process, such issues arise with constitutional safeguards in mind. The Constitution of India under Article 21 protects life and personal liberty save in accordance with procedure established by law. The requirement read into Article 21 in Maneka Gandhi is that such procedure must be just, fair and reasonable, and not arbitrary or oppressive.[10][11] Opaque algorithmic evidence may undermine this ability. When the prosecution makes use of an AI system, but is not able to disclose or explain the methodological underpinnings, then the defendant may find himself having to challenge an outcome that was reached without being able to question the process that led to this outcome. Article 14 poses another issue. An algorithmic system that is capable of creating disparate error rates may incorporate arbitrary and discriminatory outcomes into seemingly neutral technical systems.[12]

This does not imply that proprietary software used in AI should necessarily be disallowed. There are valid considerations of trade secret and intellectual property in favor of protecting commercial technology. Where those considerations clash with the accused's right to contest evidence necessary for a conviction, the courts need to find ways to resolve the tension without compromising either the confidentiality or the process of fairness. Thus, the Constitution provides an essential moral standard: the more critical the AI-generated conclusion is to the deprivation of liberty, the more substantial the reason for its secrecy must be.

[10]Maneka Gandhi v. Union of India, (1978) 1 S.C.C. 248 (India).
[11]INDIA CONST. art. 21.
[12]INDIA CONST. art. 14.

VIII. COMPARATIVE LESSONS

There is no need for Indian courts to adopt doctrines in evidence from any other country. But comparison does provide analysis that can be helpful. There is a system of careful examination of scientific evidence in American law of evidence. Testability, peer review, rate of errors, control mechanisms of standards and techniques and the like were some of the factors that emerged from Daubert v. Merrell Dow Pharmaceuticals, Inc.[13] Evidentiary AI technologies should also elicit similar queries such as: Is there a possibility of testing the conclusions reached by the system? What is the error rate of the system? Has the methodology employed been validated? Have the required technical standards been adhered to? Under what kind of conditions does the system perform well? The European Union's Artificial Intelligence Act, 2024, which classifies systems according to the risk they pose, highlights risk management, transparency and human oversight in high-risk applications of AI; India has, as yet, no comparable statute. In the broader context of evidence law, the message from the example is that reliability cannot be assessed in terms of an "either/or" determination on whether the system "employs AI" or not. The standard of legal assessment should be based on the risk involved in the use of the technology.

IX. THE PROPOSED ARTIFICIAL INTELLIGENCE EVIDENCE RELIABILITY TEST (AIERT)

This paper proposes that Indian courts employ a structured Artificial Intelligence Evidence Reliability Test whenever AI has materially generated, reconstructed or interpreted evidence.

1. Provenance

Where the data originates, how it has been collected, and how it is inputted into the AI system must be determined by the proponent. Data of uncertain provenance merits less weight.

2. Authenticity and Integrity

Courts must analyze metadata, hash codes, chain of custody, among others, to prove that the related electronic documents have not been tampered with.

[13]Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593-94 (1993).

3. Degree of AI Intervention

What exactly the AI did should be determined by the court. Simply storing or organizing evidence poses distinct dangers from creating missing pixels, synthesizing speech, or identification creation. It would be appropriate to follow the rule that the more generative the action is, the more reliability proof is needed.

4. System Reliability and Error Rate

When an artificial intelligence tool is generating an evidentiary inference, the person using it ought to provide meaningful information relating to its validation and accuracy rate.

5. Explainability

The appropriate methodology must be sufficiently intelligible to withstand legal scrutiny. Complete technical intelligibility is not possible, but the evidentiary inference is not thereby rendered proof against challenge simply because the underlying theory is complex.

6. Reproducibility

Where reasonably possible, an independent expert applying the same inputs and system configuration should be capable of reproducing or testing the result.

7. Human Oversight

The courts should decide whether there was a proper human qualification of the AI generated output, or if the acceptance was automatic. Human intervention cannot be reduced to being merely ceremonial or a mere “rubber stamp.”

8. Independent Corroboration

AI-generated evidence of high uncertainty would not normally serve as the only basis for serious outcomes, especially convictions in criminal cases. It is advisable to search for independent evidence where feasible.

9. Adversarial Challenge

The opposing side would need a fair chance to contest the AI evidence. This could be facilitated through granting access to data sets and technical literature depending on the nature of the dispute. AIERT would not be creating nine rigid statutory criteria. Rather, it would create a set of guidelines for judges to evaluate the reliability of the evidence.

X. RECOMMENDATIONS FOR THE INDIAN LEGAL SYSTEM

Firstly, there is a need for guidelines dealing specifically with the use of artificial intelligence-based evidence. Statutory definitions that are technology-specific will quickly become out of date; the focus must rather be on the extent of algorithmic processing and the attendant evidentiary risk.

Secondly, forensic laboratories and investigation agencies must develop chain-of-custody guidelines that are AI-specific. It is not enough simply to preserve the final file. In cases of AI-based processing, it is essential to keep the original input, the output file, the model/version, and any other necessary processing details.

Thirdly, there is a need for the Indian judiciary to mandate the disclosure of AI processing of evidence. No party should be allowed to produce an enhanced photo as though it was an unenhanced original. The fact of enhancement must necessarily be disclosed.

Fourthly, there is a need for courts to distinguish between AI enhancement and AI generation of data. An enhancement tool which increases brightness/contrast poses lower epistemic risks than a generative tool which creates missing information.

Fifthly, there is a need for systemic building up of technical capabilities in the judiciary. The judges need not become computer scientists themselves; however, they must know enough about topics like metadata, hashing, model confidence, false positives, training data, generative reconstruction, and algorithmic bias to be able to critically analyze the testimony of the experts.

Sixthly, it will be desirable for the courts to have independent technical experts when the systems used are proprietary or highly advanced in terms of technical complexity and are having any effect on the proceedings. Relying solely on the testimony of the experts appointed by the company behind the system or the investigating agency may not ensure the needed neutrality.

Lastly, the need for establishing the reliability of the AI should depend upon the significance of its use in the trial.

XI. CONCLUSION

AI does not make existing law of evidence irrelevant. Rather, it makes evident that existing law of evidence had been operating on certain assumptions concerning technology that no longer hold. Bharatiya Sakshya Adhiniyam sets up a necessary framework for dealing with electronic and digital evidence. The theory worked out in Anvar P.V. and Arjun Panditrao will continue to be relevant because of provenance, certification, and electronic integrity. But now AI poses a challenge not addressed by previous theories of electronic authentication. An undamaged and well-preserved electronic document can preserve an AI falsehood. Indian law of evidence needs to address the distinction between the authenticity of the electronic document and reliability of algorithmic content contained in it. Neither approach is satisfactory. Assuming AI evidence to be necessarily unreliable will deprive courts of new technologies that can advance their work. Assuming formally authenticated AI outputs to be a regular electronic document will leave courts open to deepfakes, hallucinations, hidden biases, and scientifically unverified conclusions of algorithms. A better way would be technology neutral admissibility coupled with risk sensitive reliability assessment. This could be seen from the suggested AIERT model above. The examination of provenance, integrity, level of involvement of the AI system, error rate, explainability, replicability, human supervision, corroboration and testing against adversaries will help courts assess the admissibility of AI generated evidence based on real epistemological dangers and not mere technology. Ultimately, artificial intelligence alters the question of evidence. While the courts can continue asking, "Is this the original electronic record?", they increasingly have to pose another and harder question, "How was what we see on this record generated?" The legitimacy of AI assisted adjudication process will hinge upon the ability of Indian evidence law to answer both.