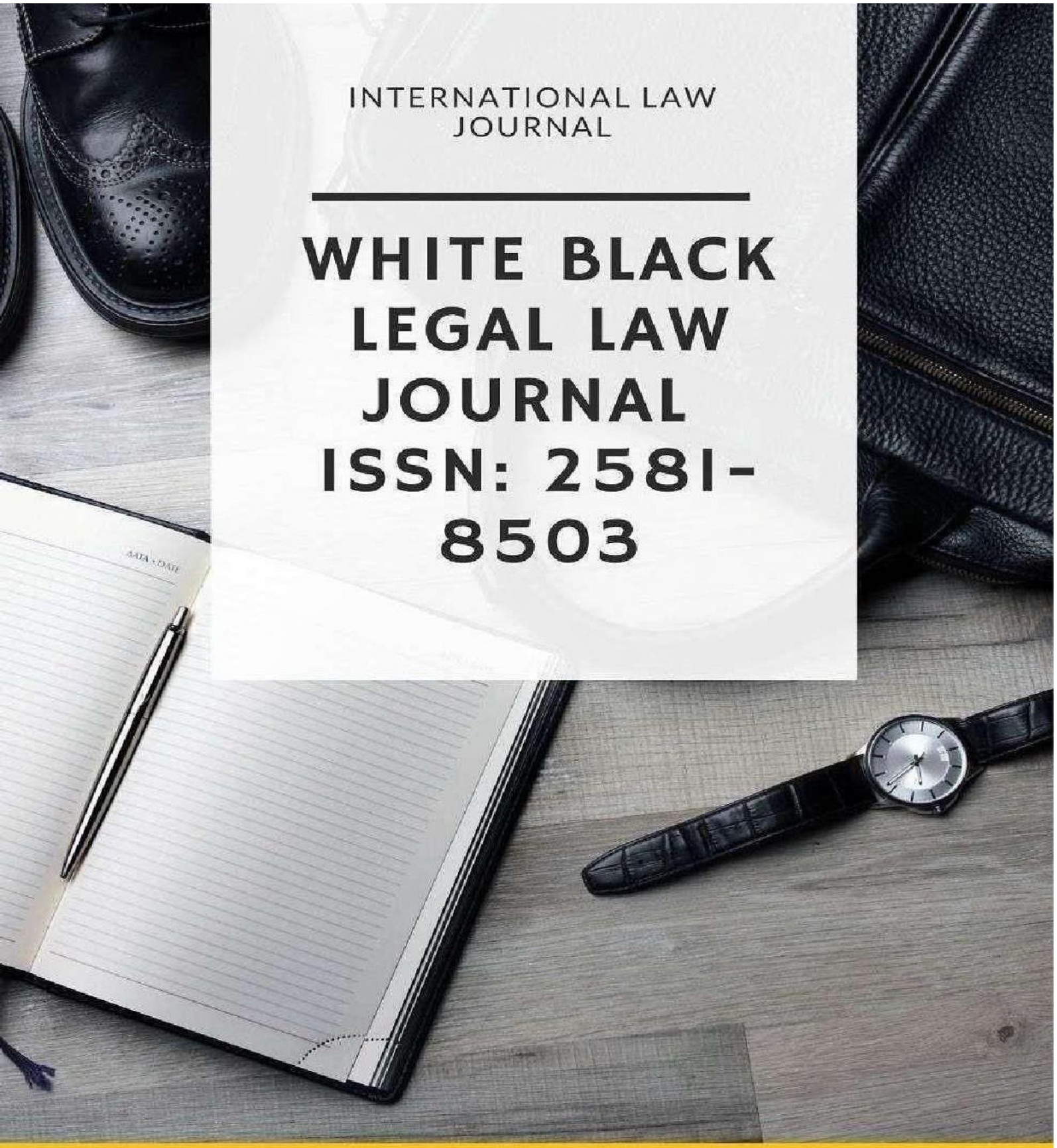


INTERNATIONAL LAW JOURNAL

WHITE BLACK LEGAL LAW JOURNAL

ISSN: 2581-8503

Peer - Reviewed & Refereed Journal

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

WWW.WHITEBLACKLEGAL.CO.IN

DISCLAIMER

No part of this publication may be reproduced, stored, transmitted, translated, or distributed in any form or by any means—whether electronic, mechanical, photocopying, recording, scanning, or otherwise—without the prior written permission of the Editor-in-Chief of *White Black Legal – The Law Journal*.

All copyrights in the articles published in this journal vest with *White Black Legal – The Law Journal*, unless otherwise expressly stated. Authors are solely responsible for the originality, authenticity, accuracy, and legality of the content submitted and published.

The views, opinions, interpretations, and conclusions expressed in the articles are exclusively those of the respective authors. They do not represent or reflect the views of the Editorial Board, Editors, Reviewers, Advisors, Publisher, or Management of *White Black Legal*.

While reasonable efforts are made to ensure academic quality and accuracy through editorial and peer-review processes, *White Black Legal* makes no representations or warranties, express or implied, regarding the completeness, accuracy, reliability, or suitability of the content published. The journal shall not be liable for any errors, omissions, inaccuracies, or consequences arising from the use, interpretation, or reliance upon the information contained in this publication.

The content published in this journal is intended solely for academic and informational purposes and shall not be construed as legal advice, professional advice, or legal opinion. *White Black Legal* expressly disclaims all liability for any loss, damage, claim, or legal consequence arising directly or indirectly from the use of any material published herein.

ABOUT WHITE BLACK LEGAL

White Black Legal – The Law Journal is an open-access, peer-reviewed, and refereed legal journal established to provide a scholarly platform for the examination and discussion of contemporary legal issues. The journal is dedicated to encouraging rigorous legal research, critical analysis, and informed academic discourse across diverse fields of law.

The journal invites contributions from law students, researchers, academicians, legal practitioners, and policy scholars. By facilitating engagement between emerging scholars and experienced legal professionals, *White Black Legal* seeks to bridge theoretical legal research with practical, institutional, and societal perspectives.

In a rapidly evolving social, economic, and technological environment, the journal endeavours to examine the changing role of law and its impact on governance, justice systems, and society. *White Black Legal* remains committed to academic integrity, ethical research practices, and the dissemination of accessible legal scholarship to a global readership.

AIM & SCOPE

The aim of *White Black Legal – The Law Journal* is to promote excellence in legal research and to provide a credible academic forum for the analysis, discussion, and advancement of contemporary legal issues. The journal encourages original, analytical, and well-researched contributions that add substantive value to legal scholarship.

The journal publishes scholarly works examining doctrinal, theoretical, empirical, and interdisciplinary perspectives of law. Submissions are welcomed from academicians, legal professionals, researchers, scholars, and students who demonstrate intellectual rigour, analytical clarity, and relevance to current legal and policy developments.

The scope of the journal includes, but is not limited to:

- Constitutional and Administrative Law
- Criminal Law and Criminal Justice
- Corporate, Commercial, and Business Laws
- Intellectual Property and Technology Law
- International Law and Human Rights
- Environmental and Sustainable Development Law
- Cyber Law, Artificial Intelligence, and Emerging Technologies
- Family Law, Labour Law, and Social Justice Studies

The journal accepts original research articles, case comments, legislative and policy analyses, book reviews, and interdisciplinary studies addressing legal issues at national and international levels. All submissions are subject to a rigorous double-blind peer-review process to ensure academic quality, originality, and relevance.

Through its publications, *White Black Legal – The Law Journal* seeks to foster critical legal thinking and contribute to the development of law as an instrument of justice, governance, and social progress, while expressly disclaiming responsibility for the application or misuse of published content.

www.whiteblacklegal.co.in
Volume 4 Issue 1 | Sep 2026 ISSN: 2581-8503

FROM AUTHENTICITY TO PROVENANCE: RECONSTRUCTING THE LAW OF DIGITAL EVIDENCE IN THE AGE OF GENERATIVE AI

AUTHORED BY - B. JAMES JAYA RAJ.

Abstract:

For two centuries, the law of evidence rested on a quiet assumption: that a photograph, a recording, or a document was, in the ordinary course, a record of reality — a trace left by an event, preserved by custody, and verifiable by authentication. Generative artificial intelligence has dismantled that assumption. Machines can now manufacture images, voices, videos, and documents that bear every visible hallmark of authenticity while depicting events that never occurred. This paper argues that in the age of generative AI, authenticity is no longer an adequate legal proxy for reliability. Digital evidence law must move from an authentication-centred model towards a provenance-centred model — one that examines not merely whether a digital artefact is what it purports to be, but how it was created, by whom, through what technological process, whether it has been altered, and whether its evidentiary meaning can be independently verified. The paper develops the Deepfake Paradox (the greater the capacity to fabricate, the easier it becomes to discredit the genuine), exposes the conceptual fallacy that authenticity equals truth, and traces the Indian framework from the Evidence Act 1872 through the Section 65B jurisprudence to the Bharatiya Sakshya Adhiniyam, 2023. It then compares the responses of the United States, the United Kingdom, the European Union, and Brazil, before proposing the instruments of reform: a Provenance Principle, a P-R-O-V-E-N-A-N-C-E judicial test, a scheme of graduated provenance standards, and a seven-part agenda of digital forensic governance. Its conclusion is deliberately provocative: the future of digital evidence belongs not to the evidence that looks real, but to the evidence whose provenance can be proved.

Keywords: deepfakes; provenance; authentication; digital evidence; Bharatiya Sakshya Adhiniyam, 2023; generative AI; evidentiary reliability; liar's dividend; synthetic media; digital forensic governance

***

I. Introduction — When Seeing Is No Longer Believing

Imagine the prosecution produces a video. It appears to show the accused committing the crime. The face is unmistakable. The voice is unmistakable. The timestamp is intact; the metadata appears genuine; a witness points at the screen and says, with quiet certainty, *"That is him."* The only problem is this: the video never happened. It was not recorded. It was generated — conjured from statistical patterns, not captured from photons. Every pixel is true to a grammar of reality that the machine learned by consuming reality, and yet the scene it depicts is a fiction.

Now reverse the situation. An accused person produces a genuine CCTV recording — indisputably captured, unbroken in custody — and the prosecution retorts: *"Anyone can manufacture such a video using AI."* No one alleges a specific manipulation. None is needed. The mere *existence* of the technology is enough to sow the doubt.

What happens when the existence of a credible fake creates doubt about the genuine?

That question is the doorway through which this paper will pass, and it deserves an answer at the threshold: *a credible fake does not merely add one more false exhibit to the record; it contaminates the epistemic environment in which genuine evidence is judged.* A legal system built around the assumption that seeing is believing must now confront the possibility that seeing is *manufacturing* (Chesney & Citron, 2018; Grossman & Grimm, 2025).

The questions multiply, and each deserves a provisional answer: *Can visual certainty still produce legal certainty?* Only if the law can verify the *origins* of the visual — which the traditional rules of authentication, as we shall see, were never designed to do.

What does "authenticity" mean when machines can manufacture apparently authentic events? Authenticity collapses from a property of the artefact into a claim about its provenance. An artefact can be perfectly "genuine" — unforged, unaltered — and perfectly false, because its very genesis was synthetic (Sharma, 2020).

Is a digital recording evidence of an event, or merely evidence of a digital artefact? The honest answer — and it is a radical one — is the latter. A recording is evidence of the event only if the chain from event to artefact can be reconstructed; absent that chain, it is evidence only of itself (Moruzzi et al., 2025).

If AI can fabricate the appearance of reality, what exactly is a court authenticating? Authentication, stripped to its bones, answers one question: "Is this what you say it is?" When machines can produce artefacts that are exactly what the proponent says they are — a video of the accused — and still be false, authentication has been rendered necessary but radically

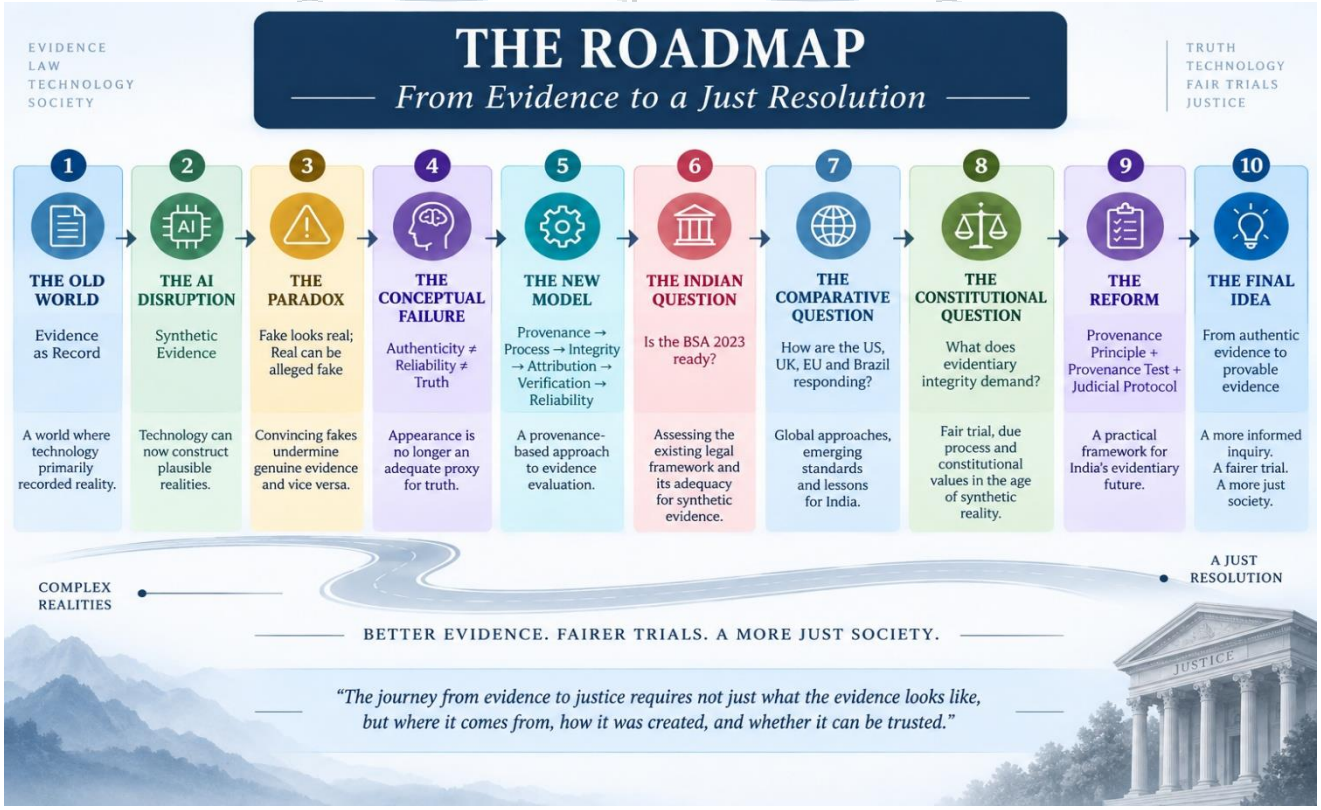
insufficient (LaMonaga, 2020).

Can a legal system built around human witnesses and physical documents adequately govern synthetic evidence? It can — but only by rebuilding its conceptual foundations. That rebuilding is the burden of this paper.

Here, then, is the first major proposition, which may stand as the paper's signature claim: *Generative AI has not merely created a new category of false evidence; it has destabilised the epistemic foundations by which courts distinguish the true from the false* (Maras & Alexandrou, 2018).

And here, promised early, is the destination: *authentication can no longer be the end of the evidentiary inquiry; it must be the beginning. The law must move from authenticating the object to proving the provenance of the object* — a principle developed in Section VII and given operative form in the P-R-O-V-E-N-A-N-C-E Test of Section XIX.

The journey of the paper can be mapped in a single glance:



II. Literature Review — What Has Been Said, and What Has Not

The literature on deepfakes and digital evidence has grown so quickly that it already stratifies into identifiable schools. It is worth mapping them, because the paper's claim to novelty can only be measured against them.

A. The Doctrinal Warning Literature

The earliest scholarly responses treated synthetic media as an existential challenge to evidence law. Chesney and Citron's foundational account identified deepfakes as a looming threat to privacy, democracy, and national security, and — with remarkable foresight — isolated the *liar's dividend* as the mechanism by which fakes corrode accountability: once audiences know that audio and video can be convincingly faked, some will "escape accountability for their actions by denouncing authentic video and audio as deep fakes" (Chesney & Citron, 2018). Maras and Alexandrou carried the warning into the courtroom, analysing how police and courts might authenticate video evidence when authenticity itself can no longer be assumed (Maras & Alexandrou, 2018). Indian scholars echoed the alarm — Sharma on the future of "evidentiary integrity" (Sharma, 2020), Goyal and Mali on the difficulties of digital evidence under the Indian legal system (Goyal & Mali, 2025), and Singh and Bamb on the structural anxiety embedded in the Section 65B certificate regime (Singh & Bamb, 2021), itself the culmination of Dubey's earlier analysis of electronic evidence admissibility (Dubey, 2017). This strand diagnoses the disease; it does not, on the whole, prescribe the cure.

B. The Empirical Deception Literature

A second strand asks what synthetic media actually *do* to human belief. Vaccari and Chadwick's experiments showed that politically targeted synthetic video produces less outright deception than *uncertainty* — and uncertainty, in turn, erodes trust in news and public communication (Vaccari & Chadwick, 2020). Twomey and his co-authors, studying public responses to deepfakes during the Russian invasion of Ukraine, found that awareness-raising interventions improved people's ability to spot fakes — while simultaneously *reducing their ability to identify genuine media by 9%* (Twomey et al., 2023). Farid's practitioner's account supplies the vivid illustration: a politician apologised for misogynistic remarks captured on tape, and a year later — knowing the technology now existed — claimed the recording was fake (Farid, 2022). Schiff, Schiff and Bueno distilled the endpoint with clinical precision: "The concern is not that 'people will be deceived, but that they will come to regard everything as deception'" (Schiff et al., 2024). This strand establishes the *psychological* mechanism of the Deepfake Paradox; it does not build the legal architecture to counter it.

C. The Technical Detection Literature

The third strand is technical: *can the machines be made to police themselves?*

Fernández Gambín and his co-authors' comprehensive survey of deepfake generation and detection methods candidly concludes that existing techniques succeed in controlled conditions while "several elements can compromise the final output" in the wild (Gambín et al., 2024). Lu and Ebrahimi's real-world assessment framework exposed the gap between laboratory promise and operational reality — compiling 204 internet videos (75 real and 129 fake) precisely because curated datasets flatter detectors, and showing that detectors trained on one type of forgery degrade sharply when confronted with another (Lu & Ebrahimi, 2024). Koutlis and Papadopoulos document why: detection models are trained on particular corpora, and their performance depends on the scale and quality of the fakes they have seen — including even the *duration* of the manipulated segments (Koutlis & Papadopoulos, 2024). And the deployment literature warns that models "evaluated solely under in-distribution assumptions may exhibit optimistic performance estimates, masking vulnerabilities that emerge during live deployment" (Rawat et al., 2026). The technical strand proves that detection cannot end the crisis; it says little about what the law should do in the meantime.

D. The Provenance-Architecture Literature

The fourth strand, newer and smaller, points away from detection toward *provenance* — building trust into the artefact at the moment of creation rather than divining it afterwards. Collomosse and Parsons articulate the "three pillars of provenance": metadata, fingerprinting, and watermarking, as the foundations of safe generative AI (Collomosse & Parsons, 2024). Kul's "digital airworthiness" protocol pairs hardware-rooted cryptographic signing at capture ("trusted capture") with resilient marking of synthetic media ("generative imprint") (KUL, 2026). Kandati goes further, proposing a deterministic, generation-time signing architecture that reduces reliance on retrospective detection altogether (Kandati, 2026). Moruzzi and her co-authors address the human dimension — how provenance data create value for creatives and audiences, and what "content authenticity" can mean in practice (Moruzzi et al., 2025). This is the most promising strand, but it remains overwhelmingly technical; it has not yet been translated into the grammar of evidence law. That translation is precisely what this paper undertakes.

E. The Comparative-Evidence Literature

The fifth strand examines how courts are, or should be, handling AI-generated evidence. Grossman and Grimm — the most important recent contribution — analyse judicial approaches to acknowledged and unacknowledged AI-generated evidence, warn that

generative AI has "democratized fraud," and conclude that fabricated content "will most certainly find their way into the resolution of court cases" (Grossman & Grimm, 2025). LaMonaga anatomised the failure of the American Rule 901 authentication standard against deepfakes (LaMonaga, 2020). Faqir compared the exclusionary treatment of AI-enhanced evidence in the United States and the UAE, exposing the unresolved validity of conclusions from machine processes (Faqir, 2024). Dhawan surveyed the admissibility frameworks of the US, EU, and UK (Dhawan, 2026), and Shamov grappled with the procedural challenges deepfakes pose to courts — including, valuably, the cost of verification and the case for burden-shifting (Shamov, 2024).

F. The Gap

Survey the field and the gap is unmistakable. The warning strand diagnoses; the empirical strand measures; the technical strand detects; the provenance strand engineers; the comparative strand describes. **None of them reconstructs the jurisprudential foundation itself — the concept of authentication on which all of it rests.** The existing scholarship asks, in a thousand idioms, whether the deepfake is a new kind of false evidence. This paper asks a different question: whether generative AI has made *authentication itself* the wrong instrument — whether the law's entire architecture of "Is this what you say it is?" is unequal to a world in which the answer can be "yes" and the evidence still false. That is the reconstructive ambition, and it is what distinguishes this paper from the growing literature on "deepfakes under the BSA."

III. Methodology and Research Design

A. Research Design

This study employs a **fourfold, mixed method**: (i) *doctrinal analysis* of Indian statutory provisions and judicial decisions governing electronic evidence, from the Indian Evidence Act, 1872 through the Section 65B era to the Bharatiya Sakshya Adhiniyam, 2023; (ii) *functional comparison* of the evidentiary responses of the United States, the United Kingdom, the European Union, and Brazil; (iii) *regulatory analysis* examining how platform and AI regulation (the EU AI Act; India's IT Rules, 2021 and 2026) can supply evidentiary infrastructure; and (iv) *conceptual-argumentative reconstruction* of the authentication doctrine itself. The fourth element is methodologically decisive: the paper's core claim — that authenticity must yield to provenance — is a jurisprudential claim, tested not by statistics but

by the coherence, completeness, and practical utility of the proposed framework.

B. Research Questions

The interrogative form is not decorative here; it is the method. The paper is driven by five questions:

1. *What exactly does authentication establish, and why did it once suffice?*
2. *What happens to the value of genuine evidence when the existence of a credible fake creates doubt about the authentic?*
3. *Are the principles developed to prove that a digital record came from a particular computer capable of establishing the provenance of an AI-generated or AI-manipulated record?*
4. *Can a machine prove another machine is lying — and who authenticates the authenticator?*
5. *Who should bear the burden of proving provenance, and with what rigour, as the consequences of the case mount?*

C. Data and Materials

The paper draws on four bodies of material. *First*, primary legal instruments: the Indian Evidence Act, 1872 (as amended by the IT Act, 2000), the Bharatiya Sakshya Adhiniyam, 2023 (Sections 61–63), the Information Technology Rules, 2021 and the Amendment Rules of 2026, the Federal Rules of Evidence 901 and 702 of the United States, and Article 50 of the EU Artificial Intelligence Act (Finck, 2026; Meding & Sorge, 2024; Veale & Borgesius, 2021). *Second*, judicial decisions: the Indian electronic-evidence line from *Anvar P.V. v. P.K. Basheer* through *Tomaso Bruno v. State of U.P.* to *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal* (Goyal & Mali, 2025; Singh & Bamb, 2021), and American authorities applying Rules 901 and 702. *Third*, the empirical and technical literature on deepfake generation and detection (Gambín et al., 2024; Koutlis & Papadopoulos, 2024; Lu & Ebrahimi, 2024; Rawat et al., 2026) and on the effects of synthetic media on belief and trust (Schiff et al., 2024; Twomey et al., 2023; Vaccari & Chadwick, 2020). *Fourth*, contemporaneous reporting of fast-moving institutional developments — the United States evidence-rule process, the D.C. Court of Appeals' criticism of AI-generated citations, Brazil's electoral-court rules, and India's 2026 amendments to the IT Rules — each identified in the text as reported and dated, since institutional responses are evolving as this paper is written.

D. Limitations

Four limitations are acknowledged candidly. *First*, the paper relies in part on news reporting of developments that may be superseded by the time of publication — the 2026 United States evidence-rule deliberations, the September 2026 D.C. Court of Appeals episode, the Brazilian electoral rules, and the 2026 IT Rules amendments are each stated as developments *at the time of writing*. *Second*, the technical literature is moving faster than any doctrinal analysis can track; the paper therefore reasons at the level of technological *capacities* (generation, manipulation, detection, provenance-marking) rather than particular systems. *Third*, the provenance model is proposed as a normative-analytic instrument; its calibration awaits the empirical record of how courts apply it. *Fourth*, the paper does not attempt to quantify the prevalence of deepfake evidence in litigation — reliable figures do not exist — and no such claim is made.

IV. The Old Epistemology of Evidence

4.1 The Physical World and Evidentiary Assumptions

Traditional evidence law worked reasonably well because it was built for a world in which reality left footprints. The canonical sequence ran**: *event → human observation → recording or document → custody → courtroom.*** At every link there was a reasonably identifiable causal chain.

A photograph presupposed a camera and a photographer. A recording presupposed a device in a particular place at a particular time. A signature presupposed a person and a pen. A document presupposed an author. A witness presupposed a human observer whose perception, memory, and sincerity could be tested by cross-examination. The evidentiary object was a *trace* — the residue of an occurrence, bearing the impress of its cause.

This is why the law could afford to be epistemically modest. It did not need to prove every link in the chain; it needed only to establish enough links to make fraud unlikely, and then to let the tribunal weigh what remained. The assumption was not that artefacts could not lie — of course they could — but that *lying required human intervention*, and human intervention left marks: a forged signature, a cut splice, a doctored negative, a paid witness (Dubey, 2017; Singh & Bamb, 2021).

4.2 Authentication as the Gateway to Trust

The traditional evidentiary logic can be stated in a single sentence: if the court can establish that the object is what the party claims it to be, the court can begin assessing its

evidentiary weight.

We must be precise, however, about the concepts that this logic bundles together, for the entire argument of this paper turns on their separation:

- **Authenticity** — Is the object genuinely what it purports to be?
- **Admissibility** — May the tribunal legally consider it?
- **Relevance** — Does it tend to prove or disprove a fact in issue?
- **Reliability** — Can the *process* by which it was generated and preserved be trusted?
- **Probative value** — How much does it actually advance the proof, all things considered?

In the old world, these five inquiries could be run in sequence with tolerable safety, because authenticity was a *strong proxy* for reliability: a real photograph was, absent human staging, a fairly dependable record of what the camera saw. The proxy worked because the causal chain was physical. What generative AI dissolves is precisely this proxy — and with it, the comfortable sequence of the old epistemology.

4.3 The Indian Position

India's statutory engagement with electronic evidence has been a long march from suspicion to accommodation. The Indian Evidence Act, 1872, with its insistence on primary documents and its deep suspicion of copies, was drafted for paper. The Information Technology Act, 2000 inserted Sections 65A and 65B, creating a special regime for the admissibility of electronic records — records whose "original" exists only as machine-readable data, and whose evidentiary avatar is necessarily a printout or a copy certified under Section 65B (Dubey, 2017; Singh & Bamb, 2021).

The jurisprudence then oscillated for two decades — the mandatory certificate questioned, relaxed, and reinstated — before the Constitution Bench in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal* settled that a Section 65B certificate was ordinarily indispensable for the admissibility of electronic records (Goyal & Mali, 2025). The four-stage journey runs: ***Indian Evidence Act, 1872 → Section 65B jurisprudence → Anvar P.V. → Arjun Panditrao Khotkar → Bharatiya Sakshya Adhiniyam, 2023*** (Goyal & Mali, 2025).

The BSA, in its Sections 61–63, now specifically recognises electronic and digital records within a modernised framework — treating them as documentary and primary evidence, and formalising their admissibility, authentication, and integrity (K & M.B, 2026; Srinivas, 2026). But here is the arresting observation: *the BSA modernises the legal recognition of digital records without confronting the generative rupture***.** Its vocabulary is the vocabulary

of the old epistemology — records, originals, custody — even as the technology it governs has begun to manufacture its own facts. Recent Indian litigation has already demonstrated the practical difficulty of transiting from the old 65B framework to the BSA's new regime, as courts wrestle with certificates, mirror images, and the sheer volume of digital material (as reported by Live Law).

The question to be answered in Section IX is not whether the BSA is an improvement — it plainly is — but whether improvement is enough.

V. The Generative AI Break — The Evidence Can Now Manufacture Itself

Somewhere between 2014 and 2024 the evidentiary universe tipped. Advances in generative models made it possible to produce synthetic text, images, audio, and audiovisual depictions so realistic that it has become nearly impossible — even for computer scientists — to tell authentic from fake content (Gambín et al., 2024). The technology is cheap, accessible, and ubiquitous; Grossman and Grimm go so far as to say that generative AI has "democratized fraud" (Grossman & Grimm, 2025). The courtroom consequence is not hypothetical: content fabricated for deception has already been used to destroy reputations, manipulate elections, and mislead publics — and it "will most certainly find their way into the resolution of court cases" (Grossman & Grimm, 2025).

It is worth mapping the new taxonomy of synthetic evidence, because each category attacks a different evidentiary assumption:

A. **Synthetic image.** A person appears somewhere they never were — in a room, at a rally, at a crime scene. The old assumption that a photograph presupposed a camera dies here.
B. **Synthetic video.** A person appears to perform an act they never performed — to strike, to confess, to conspire. Deepfake videos are the product of machine-learning applications that merge, replace, and superimpose images and clips to create a fake video that appears authentic (Farid, 2022; Maras & Alexandrou, 2018).
C. **Synthetic audio.** A person's voice appears to make a statement they never made. The ear, like the eye, is now a fallible witness.
D. **Synthetic text.** A document appears to have been written by a person who never wrote it — a contract, a confession, a will, a defamatory memorandum.
E. **Synthetic witness.** AI-generated avatars can produce apparently testimonial statements — a "witness" who did not observe, remembers nothing, and cannot be cross-examined in any meaningful sense.

F. **Synthetic reconstruction.** AI can recreate events from incomplete evidence — reconstructing a scene, an accident, or an altercation from fragments, raising the profound question of whether a reconstruction can ever be *proof* rather than *theory*.

Now the conceptual leap, stated plainly: *the courtroom is moving from a world in which technology recorded reality to a world in which technology can construct plausible realities***.** That is the fundamental transformation — and it is why every rule of evidence built on the recording assumption must now be examined brick by brick (Maras & Alexandrou, 2018; Sharma, 2020).

VI. The Deepfake Paradox — When the Fake Undermines the Real

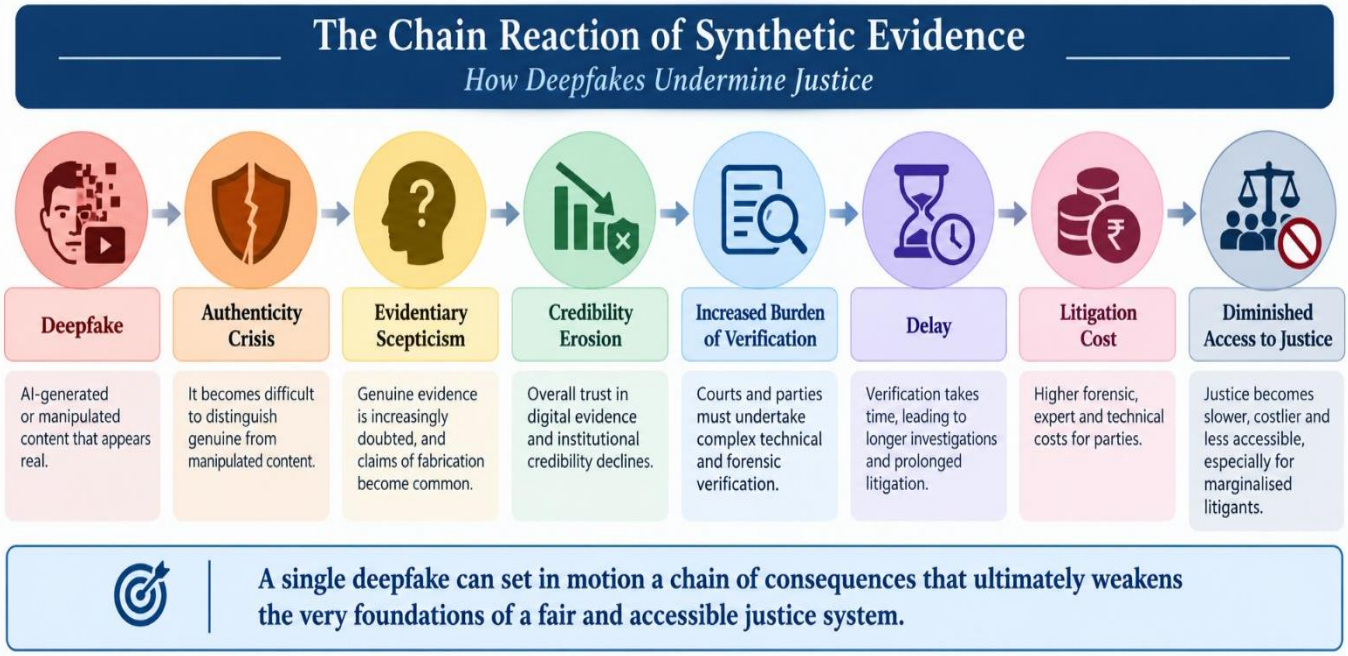
The Deepfake Paradox: The greater the ability of technology to fabricate convincing evidence, the easier it becomes to challenge genuine evidence merely by alleging fabrication. The paradox produces two symmetrical dangers, and the literature has attended disproportionately to only one:

- **False Positive** — fake evidence is *accepted* as genuine, and an innocent person is convicted, or a fraudulent claim succeeds.
- **False Negative** — genuine evidence is *rejected* as potentially fake, and a guilty person walks free, or a meritorious claim collapses.

The second danger is the neglected one — yet it is, arguably, the more corrosive, because it scales with every successive improvement in detection technology. The phenomenon is well documented outside the courtroom as the **"liar's dividend"**: as Chesney and Citron put it, once the public understands that audio and video can be convincingly faked, "some will try to escape accountability for their actions by denouncing authentic video and audio as deep fakes" (Chesney & Citron, 2018). The dividend "flows, perversely, in proportion to success in educating the public about the dangers of deep fakes" (Chesney & Citron, 2018). Farid's account is salutary: a politician apologised for misogynistic remarks captured on tape, and a year later — knowing the technology now existed — claimed the recording was fake (Farid, 2022). Schiff and her co-authors crystallise the endpoint: "The concern is not that 'people will be deceived, but that they will come to regard everything as deception'" (Schiff et al., 2024).

The experimental evidence is even more unsettling, and it deserves to be the empirical heart of this section. Interventions designed to help people spot deepfakes measurably improved their ability to identify fakes — while simultaneously **reducing their ability to correctly identify real media by 9%** (Twomey et al., 2023). In other words, awareness of the threat degrades trust in the genuine; the antidote carries its own poison.

The courtroom translation of the paradox runs as a descending chain:



Deepfakes, on this showing, are not merely an *evidence* problem. They are, in the most literal sense, a justice-system efficiency problem — one that taxes every litigant, genuine or fraudulent, and that falls hardest on those least able to marshal the forensic firepower to rebut an innuendo of fabrication (Pawelec, 2022; Tortora, 2024). This is where the paradox acquires its constitutional bite: the liar's dividend most cruelly burdens those "society is already less likely to believe" (Schiff et al., 2024) — the marginalised defendant, the low-income complainant, the impecunious party whose genuine recording is dismissed with a shrug and the words "anyone can fake that." For them, the cost of verification is not an inconvenience but an insurmountable barrier; and a procedural system that lets doubt flourish wherever money cannot answer it is a system that has quietly rationed justice by wealth (Shamov, 2024). The Deepfake Paradox is therefore not only an epistemic problem — it is a distributive one.

VII. From Authenticity to Provenance — The Paper's Core Contribution

The diagnosis carries within it the prescription.

7.1 Authentication asks: *"Is this what you say it is?"*

7.2 Provenance asks: *"Where did this come from?"* — and, pushed further, "how was it made, by whom or by what, and what has happened to it since?"

These are not the same question, and the difference is not stylistic. Authentication is an *identity* inquiry about the artefact; provenance is a *biography* inquiry about its entire existence. The one looks at the object; the other reconstructs its life.

A properly conceived provenance inquiry should therefore ask, at minimum:

- **Origin** — Where was the evidence created?
- **Creator** — Who or what created it — a human, a device, a model?
- **Process** — What technological process produced it — capture, generation, editing, restoration?
- **Integrity** — Has it been altered, and can any alteration be detected and dated?
- **Chain of custody** — Who possessed it, and with what breaks?
- **Transformation** — Has AI processed it — enhanced, denoised, upscaled, in-painted — at any stage?
- **Attribution** — Can the content reliably be attributed to the person depicted or heard?
- **Verification** — Can independent technical methods corroborate its claims about itself?
- **Context** — Has the content been extracted from a larger context that changes its meaning?
- **Purpose** — Why and when was it created, and for what audience?

This is not a merely academic list. It tracks, in legal form, the shift already underway in the technical literature, where the insufficiency of probabilistic detection has driven scholars toward provenance architectures: metadata, fingerprinting, and watermarking as "the three pillars of provenance" (Collomosse & Parsons, 2024); hardware-rooted cryptographic signing at the moment of capture ("trusted capture") coupled with resilient marking of synthetic media ("generative imprint") (KUL, 2026); and deterministic, generation-time signing frameworks that reduce reliance on retrospective detection entirely (Kandati, 2026). The law has not yet caught up with this technical turn — and the gap between the two is the seam this paper sews.

Hence the formulation of the paper's strongest original contribution — the **Provenance Principle**: ***No synthetic or digitally mediated evidence should be evaluated solely by asking whether it appears authentic; courts should examine the evidence's provenance, technological history, integrity, attribution and independent verifiability.***

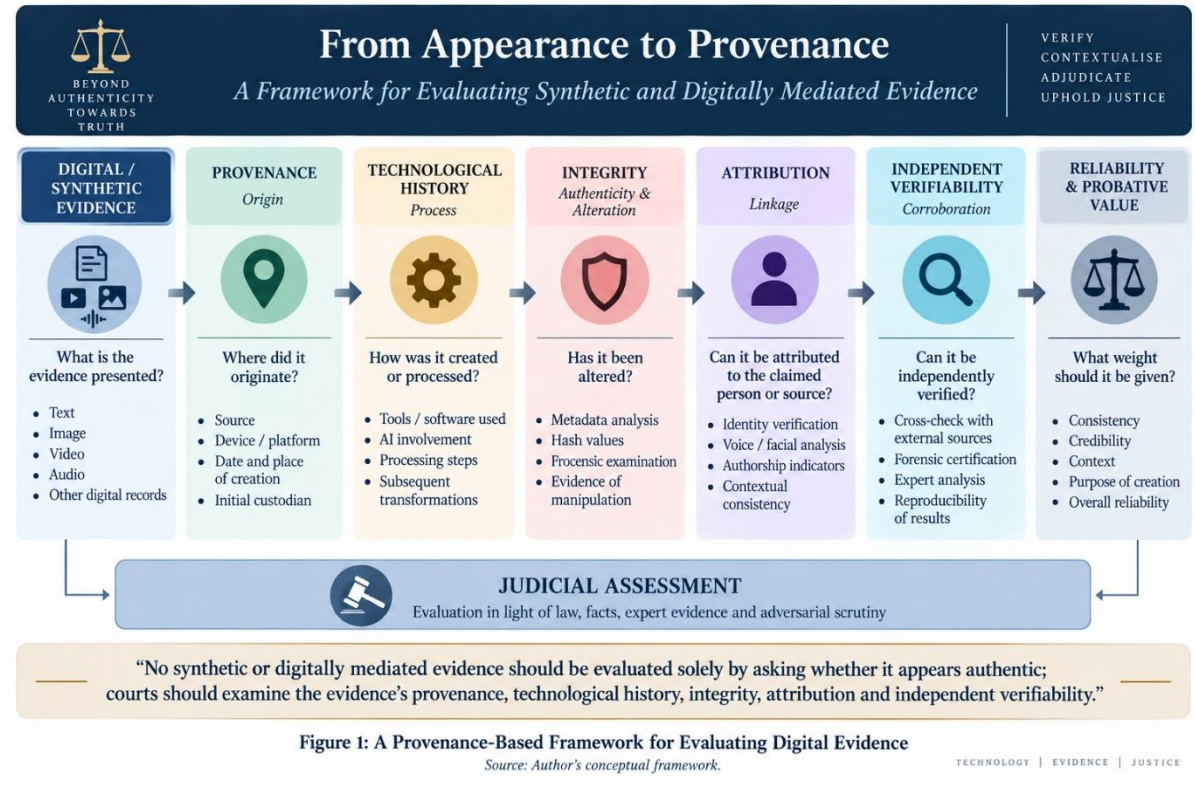


Figure 1: A Provenance-Based Framework for Evaluating Digital Evidence
Source: Author's conceptual framework.

VIII. Authenticity ≠ Reliability ≠ Truth

It is now possible to give the argument its jurisprudential spine. The law has habitually treated these three terms as a graduated ladder — if the artefact is authentic, it is likely reliable, and therefore probably true. Generative AI breaks every rung.

Layer 1 — Authenticity. Is the digital object genuine — unforged, unaltered, what it claims to be? This is the layer the rules of authentication police, and it is the one generative AI renders cheapest to satisfy. A deepfake is "authentic" as a file: it is exactly what it appears to be — a video of a person doing a thing. Its authenticity is immaculate; its truth is nil.

Layer 2 — Reliability. Can the *process* by which the object was generated or preserved be trusted? Here the inquiry shifts from the object to its biography: the capture device, the generative model, the edit history, the custody. Reliability is a process concept, and it is the layer at which provenance must operate.

Layer 3 — Truth. Does the evidence actually establish the proposition for which it is offered? This is the terminal inquiry, the one for which the earlier layers are merely scaffolding.

The three layers pull apart in directions that conventional deepfake scholarship has missed, and

we should savour the paradox:

A genuine digital file can contain false information. A real photograph can capture a staged event; an unaltered audio file can record a lie or a misleading fragment. Conversely, **a synthetically generated file may accurately reconstruct an event.** An AI reconstruction may depict, with rough fidelity, an event that actually occurred but was never recorded — a reconstruction that is "not original" yet not "false". Therefore: **"Not original" does not necessarily mean "false"; and "original" does not necessarily mean "true."**

This disjunction is the paper's escape from the trap of writing yet another article that simply declares the BSA inadequate for deepfakes. The more sophisticated thesis — and the one that gives this paper its reason for being — is that **the law must stop treating the digital artefact itself as the endpoint of authentication.** The evidentiary inquiry must move backwards: from the artefact to its provenance, from provenance to process, and from process to reliability.

IX. The Indian Evidentiary Framework — Is the BSA Ready?

The question now presses with full force upon Indian law. The Bharatiya Sakshya Adhiniyam, 2023, replaces the Indian Evidence Act, 1872, and its treatment of electronic records is, by any measure, a modernisation. Sections 61–63 of the BSA recognise electronic and digital records as admissible evidence, treat them as documentary and even primary evidence, and streamline the certificate regime that the 65B jurisprudence had made labyrinthine (Goyal & Mali, 2025; Singh & Bamb, 2021; Srinivas, 2026).

The statutory scheme now expressly contemplates: the admissibility and proof of electronic records; certificates of authenticity; expert evidence on technical questions; and — through the companion criminal-law reforms — a digitally native process in which forensic evidence, metadata, and hash values are routine currency (K & M.B, 2026). The BNSS permits electronic FIRs, audio-video recording of investigative procedures, and virtual testimony; the BNS recognises technology-driven offences. India has, in short, committed itself to a digital adjudication ecosystem (Srinivas, 2026).

Yet the question this section exists to ask must now be put with full force: **the BSA's conceptual vocabulary remains the vocabulary of the old epistemology.** Its certificates authenticate *that a record came from a particular device or process*; they do not, on their face, authenticate *the genealogy of the content itself* — whether the record was generated, manipulated, or reconstructed by an AI system before it ever reached the device in question

(Goyal & Mali, 2025; Singh & Bamb, 2021).

Recall the line of authority. *Anvar P.V. v. P.K. Basheer* insisted on the Section 65B certificate as the exclusive route for secondary electronic evidence; the jurisprudence then wobbled through *Shafhi Muhammad* before the Constitution Bench in *Arjun Panditrao Khotkar* restored the certificate's centrality (Goyal & Mali, 2025). And between them stands *Tomaso Bruno v. State of U.P.*, which deserves closer attention than it usually receives: there, the Allahabad High Court engaged with the *content* of digital evidence — the authentication of a mobile-phone video through its technical characteristics, expert examination, and the integrity of the device — and in doing so gestured, almost without noticing, at the distinction this paper makes central. *Tomaso Bruno* is significant precisely because it straddles the two eras: it certified a *device*, while the question that now haunts every courtroom is how to certify a *generation*.

The doctrinal achievement of this line was to discipline *custody and device integrity*. But consider, honestly, what it does not do.

Suppose a video is produced, a certificate is filed, the device is identified, the chain of custody is immaculate. Every question the old framework knows how to ask has been asked and answered. Now pose the new question: *the content itself — the face, the voice, the scene — was it generated by a model rather than captured by a sensor?* The certificate does not answer this; the chain of custody cannot answer it; the metadata can be forged or stripped; and the device itself may be innocent — the mere container of a manufactured artefact.

Would the principles developed for proving that a digital record came from a particular computer adequately establish the provenance of an AI-generated or AI-manipulated record? The answer, developed at length in this paper, is no. The BSA has modernised the *form* of digital evidence; generative AI has changed the *nature* of the thing being evidenced. Form and nature are now out of joint (Goyal & Mali, 2025; Grossman & Grimm, 2025).

X. The Problem of the AI Detector — Can a Machine Prove Another Machine Is Lying?

The intuitive answer to the authentication crisis is technology itself: if machines can fabricate, let machines detect. But the intuition founders on a second-order difficulty that deserves its own treatment, because here the evidentiary problem becomes recursive.

Suppose AI creates a deepfake. Another AI detects the deepfake. The court receives the

detector's conclusion. And then the question detonates: **who authenticates the authenticator?**

The chain runs: **AI-generated evidence → AI detection → expert interpretation → judicial assessment.** Each link is a potential point of failure, and the failures are structural, not incidental.

What if the detector produces a false positive? Genuine content flagged as fake is precisely the mechanism by which the liar's dividend operates on authentic evidence (Schiff et al., 2024; Twomey et al., 2023). *What if it produces a false negative?* A fabricated artefact slips through to convict. Either error is a quiet catastrophe.

What if different detectors disagree? Detection models are trained on particular corpora, and their performance is a function of their training data — its scale, the quality of the fakes it contains, even the duration of the manipulated segments (Koutlis & Papadopoulos, 2024). A detector trained on easier fakes overfits and generalises poorly to harder, partially manipulated material. There is no canonical detector, and hence no canonical answer.

What if the underlying model is proprietary? What if its training data are undisclosed? What if its error rate is unknown? These are not hypothetical anxieties; they are the ordinary conditions of commercial deployment. The technical literature is candid about the gap between laboratory promise and real-world performance: "models evaluated solely under in-distribution assumptions may exhibit optimistic performance estimates, masking vulnerabilities that emerge during live deployment" (Rawat et al., 2026). In-the-wild testing confirms the fragility — Lu and Ebrahimi compiled 204 internet videos, 75 real and 129 fake, precisely because curated datasets flatter detectors (Lu & Ebrahimi, 2024). Cross-manipulation experiments show that a model trained on one type of forgery degrades sharply on another (Lu & Ebrahimi, 2024). The research consensus is blunt: existing techniques succeed in controlled conditions, while "several elements can compromise the final output" in practice (Gambín et al., 2024).

And what of the deeper problem? Detection is a moving target. Methods that detect the fakes of today are trained on the fakes of yesterday; as generators improve, detectors decay — a race in which the defence is perpetually a generation behind (Gambín et al., 2024; Maras & Alexandrou, 2018).

The conclusion therefore deserves emphatic statement: **AI detection cannot become a new evidentiary black box replacing the old one.** A detector's output is itself data requiring provenance — its model, its validation, its error rates, its confidence intervals. The machine, too, must prove its own biography.

XI. Comparative Jurisprudence I — The United States

The Indian predicament is not unique; it is the global predicament, and the comparative record is instructive precisely because even the most mature evidence systems are faltering.

The American framework rests on Federal Rule of Evidence 901, which requires the proponent to "produce evidence sufficient to support a finding that the item is what the proponent claims it is," and on Rule 901(b), which permits authentication through the testimony of "a witness with knowledge." In practice, courts admit photographic and video evidence through the **"fair and accurate portrayal"** standard — a famously low bar. As LaMonaga observes, the witness need only testify that the depiction fairly and accurately portrays what she saw (LaMonaga, 2020).

The system worked, historically, for a contingent reason: courts relied on expert witnesses to root out fraudulent evidence — a fake that passed the low authentication bar would later be debunked at trial (LaMonaga, 2020). Deepfakes sever that assumption. Witnesses cannot identify subtle but important alterations, experts cannot reliably tell real from fake, and so fraudulent evidence, authenticated through the Rule 901(b) standard, will increasingly enter courtrooms with a decreasing ability for witnesses and courts to identify fakes (LaMonaga, 2020).

Expert evidence itself offers no refuge. Rule 702 and the *Daubert* gatekeeping factors — testing, peer review, known error rates, general acceptance — were designed for scientific methods whose operation is transparent and reproducible. Applied to machine-generated evidence, they founder: there are limited studies on commercial automated systems, and the validity of conclusions from machine processes remains an open question (Faqir, 2024).

The institutional response has been appropriately anxious. The U.S. Judicial Conference's Advisory Committee on Evidence Rules has been actively considering whether existing rules suffice for AI-generated evidence and deepfakes — and, as reported by Reuters, consideration of proposed amendments was postponed in 2026 while further expert input is sought. Even more telling, in September 2026 the D.C. Court of Appeals sharply criticised lawyers who had relied on nonexistent, AI-generated legal authorities (as reported by Reuters). Both episodes — each a development at the time of writing — converge on a conclusion of enormous comparative significance*: even a mature evidence system such as the Federal Rules is discovering that conventional authentication and expert-evidence principles may require recalibration* (Grossman & Grimm, 2025; LaMonaga, 2020).

XII. Comparative Jurisprudence II — The United Kingdom: From Digital Authenticity to Digital Integrity

The common-law tradition, in its English home, has taken a characteristically pragmatic path: no statutory code of electronic evidence, but a flexible law of authentication, expert evidence, and disclosure, periodically supplemented by guidance on digital material. The doctrinal question that divides the approaches is sharp and directly relevant to our thesis: *should a party merely demonstrate that a file has not been altered after acquisition — or must the party establish how the underlying content itself came into existence?* (Dhawan, 2026)

The first formulation is an *integrity* test: it verifies the stability of the artefact from seizure to trial. The second is a *provenance* test: it verifies the genealogy of the content — its creation, its generation, its manipulation history. The English tradition has been comfortable with the first and shy of the second, reflecting the old epistemology's assumption that content reliably descends from capture. Comparative scholarship on the UK position notes exactly this structure: judicial discretion and common-law latitude, with emerging guidance on expert and digital evidence rather than legislative prescription (Dhawan, 2026). The UK thus exemplifies the halfway house: digital integrity acknowledged, content provenance unmet — and the gap is precisely where the synthetic artefact slips through.

XIII. Comparative Jurisprudence III — The European Union: Trust Through Traceability

The European Union has chosen the most systematic route yet attempted: regulation by traceability. The EU Artificial Intelligence Act imposes, through its transparency chapter, obligations of direct evidentiary significance. Under Article 50, providers of AI systems that generate synthetic audio, image, video, or text must ensure outputs "are marked in a machine-readable format and detectable as artificially generated or manipulated," while deployers face disclosure obligations regarding manipulated content (Meding & Sorge, 2024). The draft iterations of what became the deepfake disclosure obligation required users to "disclose the artificial nature of the resulting content" where it would falsely appear authentic (Veale & Borgesius, 2021). The Act — whose implications for research and media Helberger and Diakopoulos were among the first to map (Helberger & Diakopoulos, 2022) — thus converts a regulatory duty into something the law of evidence can exploit: a *label of origin* (Finck, 2026; Meding & Sorge, 2024).

This is the paper's first original bridge between regulation and evidence: *a regulatory*

requirement to label synthetic content can become an evidentiary infrastructure. A marked file arrives in court with its genealogy pre-printed; an unmarked file that should have been marked arrives with a presumption-engendering silence. The EU approach is not flawless — it relies on the good faith of creators and does not, by itself, defeat malicious falsification of evidence (Meding & Sorge, 2024; Shamov, 2024) — but it demonstrates that provenance can be institutionalised rather than merely urged upon courts.

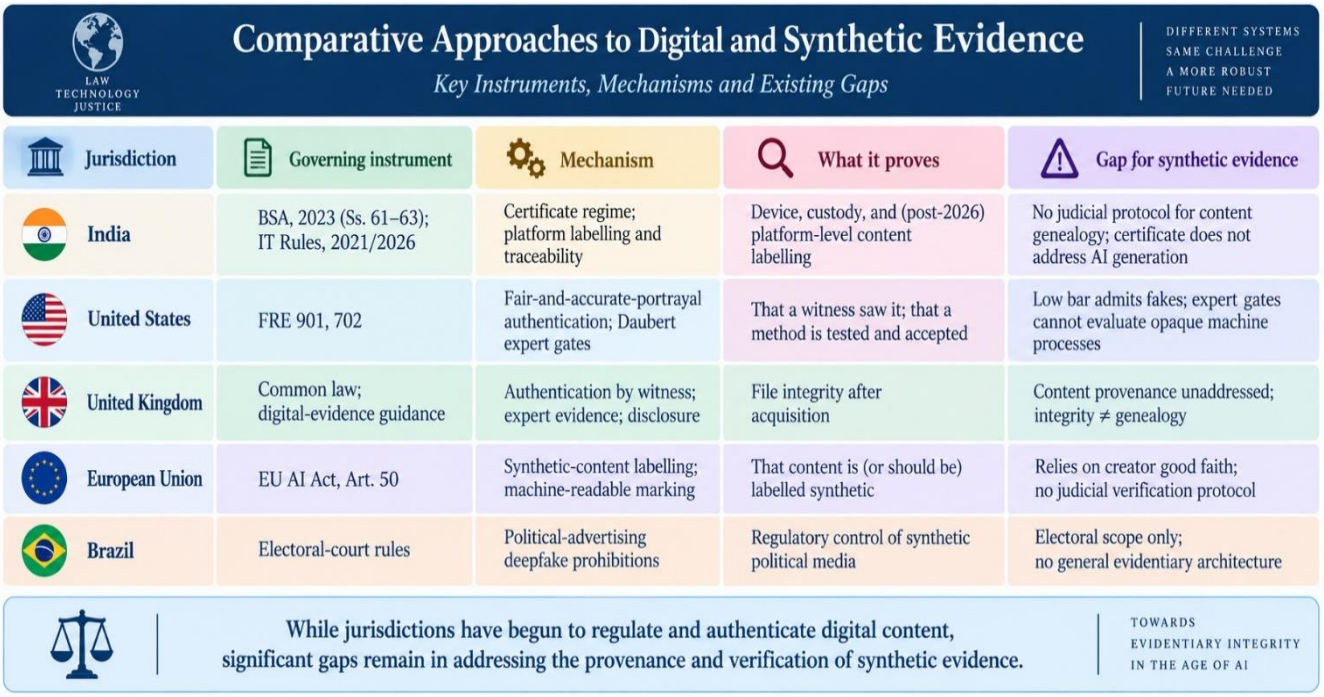
XIV. Comparative Jurisprudence IV — Brazil and the Political Evidence Problem

The courtroom, however, is not the only arena where synthetic evidence decides outcomes. Brazil supplies the contemporary demonstration: the country's top electoral court has developed rules governing AI-generated deepfakes in political advertising ahead of the 2026 elections (as reported by AP News — a development at the time of writing).

The Brazilian development broadens the paper's aperture decisively. Deepfakes now influence *elections, public perception, criminal investigations, journalism, employment disputes, family litigation, reputation, and constitutional rights* (Pawelec, 2022; Tortora, 2024). The epistemic crisis of synthetic media is thus, at one and the same time, a legal crisis and a democratic crisis. Research on political deepfakes confirms the mechanism: synthetic video induces less deception than *uncertainty* — and uncertainty, in turn, erodes trust in public communication (Vaccari & Chadwick, 2020). The threat to democracy may be most acute precisely when the fake does not deceive anyone: the *suspicion* of fabrication corrodes trust in the integrity of elections themselves (Chesney & Citron, 2018). When even authentic recordings can be dismissed as manufactured, the electorate loses the capacity to hold power to account — the liar's dividend, now playing at the scale of the *demos* itself.

XV. The Comparative Table — What No Jurisdiction Yet Proves

The comparative survey yields an observation that is more important than any single jurisdiction's rule. Set side by side, the systems diverge in mechanism but converge in limitation:



No jurisdiction — the table makes the point inescapable — has yet built a judicial instrument that *proves the genealogy of content*: where it came from, how it was made, and whether AI touched it. The provenance gap is universal, and the P-R-O-V-E-N-A-N-C-E Test of Section XIX is offered, candidly, as a contribution toward filling it.

XVI. India's New Regulatory Environment — Provenance on the Platforms

India, meanwhile, has begun to build provenance obligations at the level of platform regulation. The Information Technology Rules, 2021, established the base layer: eleven categories of prohibited content, mandated takedown of flagged content within 36 hours, a 24-hour window for removing content involving nudity or deepfakes, and traceability mechanisms empowering identification of the first originator of flagged content (Nitol, 2026).

The Amendment Rules of 2026 go considerably further (as reported by Live Law, and analysed in the literature) (GAUTAM, 2026). They formally define **Synthetically Generated Information**; compress the government-directed takedown window from 36 hours to *three*; require platforms to act within *two* hours on high-sensitivity violations; mandate **AI labelling, persistent metadata for traceability, and pre-publication verification** in which users disclose synthetic content before it is published; and, crucially, tie compliance to the Section 79 safe-harbour protection — an intermediary that fails to apply the precautions forfeits its legal immunity (GAUTAM, 2026).

The architecture is remarkable, and its logic is precisely the logic of provenance: **label the synthetic at birth; make the label travel with the artefact; hold the platforms accountable for the label's absence.** Yet the evidentiary translation remains incomplete. What the Rules create is *content provenance* — a regulatory trail for platforms. What the BSA has not yet created is *evidentiary provenance* — a judicial protocol for using that trail in court (Singh, 2026).

Here, then, is one of the paper's strongest conclusions, and it deserves to be stated with full force: **India is beginning to build provenance obligations in platform regulation, but has yet to fully translate the idea of provenance into its law of evidence.** The regulatory pipe is laid; the evidentiary tap has not been turned.

XVII. Constitutionalising Digital Truth

The argument now ascends from the technical to the constitutional, because the stakes of evidentiary failure in India are constitutional in the most concrete sense.

Article 14 guarantees non-arbitrary adjudication. A procedure that cannot distinguish synthetic from genuine evidence — that convicts on the strength of a manufactured artefact, or acquits on the strength of a manufactured doubt — is arbitrary in its operation and unequal in its incidence, for the burden of evidentiary scepticism falls heaviest upon those with the fewest forensic resources (Schiff et al., 2024).

Article 19 protects speech and expression, and with it the integrity of public information. The right to form opinions presupposes an information environment in which the genuine can be distinguished from the manufactured; deepfake-mediated disinformation degrades the very substrate of democratic deliberation (Pawelec, 2022; Vaccari & Chadwick, 2020).

Article 21 guarantees life, liberty, and fair procedure. It is here that the evidentiary crisis becomes a liberty crisis in both directions at once: a person should not be convicted because a synthetic artefact was mistaken for genuine evidence — but equally, **a genuine digital record should not become unusable merely because technology makes fabrication theoretically possible.** Both eventualities are denials of fair process; both are enabled by the same epistemic collapse.

From these premises emerges the paper's cautious constitutional claim — the **Constitutional Right to Evidentiary Integrity**: constitutional due process requires courts to employ procedures capable of distinguishing reliable digital evidence from synthetic

manipulation. This is not proposed as a new fundamental right in the doctrinal sense — Indian constitutional law does not manufacture rights by judicial fiat — but as a **constitutional principle informing evidentiary procedure**: a canon of construction that directs how admissibility rules and their implementing protocols must be read and applied (Shamov, 2024). If a defendant's liberty may turn on a pixel, the Constitution obliges the state to ensure that the pixel can be interrogated.

XVIII. The New Burden of Proof — Who Must Prove What?

The constitutional principle raises an immediate procedural question, and it must be answered with candour: if the accused says *"this video is a deepfake,"* who bears the burden? If the prosecution says *"this video is genuine,"* what level of technical verification is required?

The temptations are symmetrical, and both must be resisted. To place the burden of proving genuineness on the proponent *in every case* would drown the system in forensic certificates and price the ordinary litigant out of justice. To place the burden of proving fabrication on the challenger *in every case* would hand the liar's dividend to every defendant and every respondent with a plausible pretext (Schiff et al., 2024; Twomey et al., 2023).

The resolution lies in differentiating consequences — a principle the law of evidence already honours, since the standard of proof, and the rigour of scrutiny, vary with the gravity of what is at stake. This paper therefore proposes **Graduated Provenance Standards**:

- **Civil disputes** → *ordinary reliability*: the artefact is admitted upon reasonable authentication, with provenance questions going to weight.
- **Serious criminal prosecutions** → *enhanced provenance*: the court should require the proponent to show the artefact's origin, the process of its production or generation, and its custody — failing which, adverse inferences or exclusion.
- **Capital or maximum-penalty consequences** → *exceptionally rigorous verification*: independent forensic examination, disclosure of AI involvement, and a demonstrated absence of reasonable doubt as to non-manipulation.

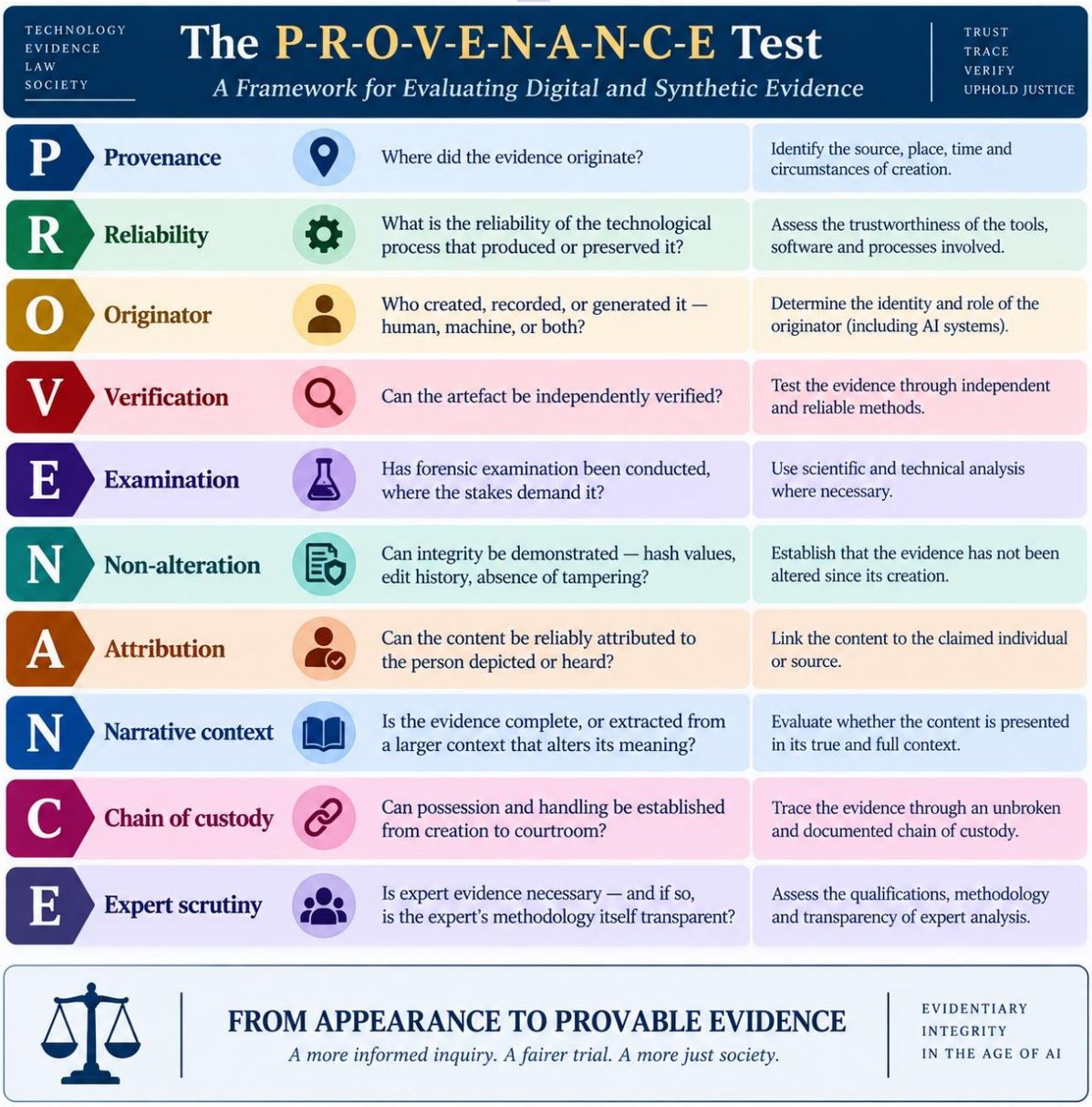
The animating idea is proportionality: **the more serious the consequence, the stronger the provenance requirement.** The standard of proof already scales with consequence; the *standard of verification* should scale with it too.

Equally important is the allocation of forensic cost. Shamov's proposal — a burden-shifting framework activated only when the challenging party provides a good-faith basis for doubt, coupled with mechanisms for allocating the high costs of digital forensics (Shamov,

2024) — offers the most sensible template. It prevents bare allegations from paralysing genuine evidence (the false-negative danger) while ensuring that serious doubt triggers serious investigation (the false-positive danger). And because the cost of verification is the hidden tax of the Deepfake Paradox, the framework must include court-appointed technical assistance where resources are unequal — otherwise the graduated standard, however principled, operates as a toll booth on the road to justice, paid only by the poor.

XIX. A Proposed Indian Framework — The P-R-O-V-E-N-A-N-C-E Test

It remains to convert principle into protocol. This paper proposes a judicial tool for synthetic and digitally mediated evidence, designed to be memorable enough for use, and named for what it does:



The Test is deliberately cumulative rather than threshold-based: a court works through the factors appropriate to the case, and the *weight* of the evidence — rather than a rigid admissibility gate — is calibrated to how many factors are satisfied and how well. In serious cases, the P-R-O-V-E-N-A-N-C-E inquiry should be conducted at a preliminary hearing, in the model of the judge as an active gatekeeper deciding authenticity before the trier of fact is exposed to the exhibit — a role the comparative literature increasingly commends (Grossman & Grimm, 2025; LaMonaga, 2020).

XX. Should AI-Generated Evidence Be Presumptively Inadmissible?

A candid scholar must confront the exclusionary possibility head-on. The debate is real, and both positions have genuine force.

The argument for exclusion runs: manipulation risk; opacity of generative processes; unknown provenance; the spectre of wrongful conviction on manufactured artefacts; and technological asymmetry between parties — one side with forensic resources, the other without (Faqir, 2024). If we cannot yet reliably tell the synthetic from the genuine, prudence might dictate that neither should be believed.

The argument against exclusion runs with equal force: AI may reconstruct genuine events that were never recorded; AI-assisted forensic analysis can be valuable; technological origin alone does not establish falsity — a synthetic reconstruction can be true in substance; and exclusion could deprive courts of the very evidence most needed in an age when the most serious crimes are documented, if at all, digitally (Grossman & Grimm, 2025).

The paper's middle position is the sophisticated one, and it should be stated with precision: **The law should not create a presumption against AI-generated evidence merely because it is synthetic; it should create a presumption against unexplained synthetic evidence.**

Synthetic origin is a *fact about provenance* — not a verdict on truth. A labelled reconstruction, transparently generated and independently corroborated, may carry genuine probative value. An unlabelled, unexplained artefact whose provenance cannot be reconstructed is in an entirely different epistemic category. The first deserves admission with calibrated weight; the second deserves suspicion, and in serious cases, exclusion. The distinction between *synthetic* and *unexplained* is the entire difference (Collomosse & Parsons, 2024; Grossman & Grimm, 2025).

XXI. The Judge in the Age of Synthetic Reality

What becomes of the judicial role when the ground itself shifts? Traditionally, the judge asks: *"Is this evidence admissible?"* Tomorrow, the judge may need to ask: *"What technological process produced this evidence? What are its known failure modes? What does the forensic expert actually establish? Can the opposing party meaningfully challenge the model?"* — questions of a different order, requiring a different competence.

The honest answer to the question *must judges become programmers?* is no. **Judges must become technologically competent adjudicators — not engineers.** The distinction is vital: a judge need not build a deepfake detector, but must be able to interrogate one; need not understand a generative model's architecture, but must know what questions to ask of it. The comparative record confirms that the judicial system is as yet unprepared for evidence derived from AI systems, and that its unpreparedness compounds complexity and cost — creating a demand for technical experts knowledgeable in generative AI and deepfake technologies, further increasing the cost and duration of legal proceedings (Grossman & Grimm, 2025; Tortora, 2024).

The institutional agenda follows directly:

- **Judicial AI literacy** — training as a core competence, not an elective;
- **Forensic training** — exposure to the methods, limits, and error rates of digital forensics;
- **Accredited digital-evidence experts** — a register of qualified examiners with disclosed methodologies;
- **Standardised forensic protocols** — uniform procedures for handling, hashing, and testing electronic exhibits;
- **Judicial technology cells** — in-house technical capacity within High Courts and the Supreme Court;
- **Independent technical assessors** — court-appointed neutral experts where parties' resources are asymmetric (Shamov, 2024).

XXII. The Deeper Jurisprudential Question — What Is Truth in the Digital Courtroom?

At the summit of the argument stands a philosophical question that the doctrinal machinery cannot evade. The law of evidence is, at its core, a truth-seeking enterprise —

Bentham's science of fact-finding, Wigmore's science of proof, the adversarial testing of testimony, the probabilistic reasoning of the trier of fact. What happens to that enterprise when the raw material of proof can be manufactured?

Pose the question in its starkest form: **If a machine can produce an image of an event that never occurred, has the image become false evidence — or is it evidence of a fabricated representation?** The answer matters, because it determines the frame: if we treat the image as a false record of an event, we ask — fruitlessly — whether the event occurred. If we treat it as a record of a *fabrication*, we ask the productive question — who made it, why, and with what intent? The second frame is the provenance frame, and it is the only one that makes progress.

Pose the second question: **If a machine reconstructs an event that actually occurred, but no camera ever recorded it, can synthetic reconstruction ever contribute to proof of the event?** Here the conventional categories fail us entirely, and it is worth being honest about why. The reconstruction is not a record; it is a hypothesis rendered in pixels. It cannot *prove* the event occurred — but it can illustrate a theory, test a hypothesis against known constraints, or corroborate oral testimony about it. Its evidentiary role is not documentary but demonstrative; its value lies not in its authenticity but in the transparency and validity of the process that produced it (Grossman & Grimm, 2025).

Both questions converge on the same doctrinal reform, and the reform deserves statement as the paper's conceptual heart: **The law should stop treating the digital artefact itself as the endpoint of authentication. The evidentiary inquiry must move backwards — from the artefact to its provenance, from provenance to process, and from process to reliability.**

This is the reconstructive ambition of the paper: not the rejection of digital evidence, but the reconstruction of the epistemic path by which digital evidence earns belief.

XXIII. Way Forward for India — From Digital Evidence to Digital Forensic Governance

The recommendations follow from the analysis, and they are offered neither as lament nor as aspiration, but as architecture — seven reforms, each traceable to a demonstrated deficit:

1. **Statutory recognition of provenance.** The BSA should expressly recognise provenance as an evidentiary consideration — by amendment or by practice direction — importing the ten-factor inquiry into the statutory text. The modernisation of 2023

that recognised electronic records (Goyal & Mali, 2025; Srinivas, 2026) must be followed by a second modernisation that recognises the *biographies* of those records.

2. **Synthetic-content disclosure.** Parties should be obliged to disclose whether evidence has been **AI-generated, AI-enhanced, AI-restored, AI-transcribed, or AI-reconstructed**. The distinction between acknowledged and unacknowledged AI evidence (Grossman & Grimm, 2025) should become a formal Indian disclosure category. A party that fails to disclose what it knows is visited with inferences; a party that discloses gains the benefit of transparency.
3. **Provenance certificates.** India should move beyond the Section 65B/BSA authenticity certificate — which certifies device and process — toward a **digital provenance certificate** that certifies content genealogy: creation method, generative involvement, edit history, and integrity checks (Singh & Bamb, 2021). The certificate is the evidentiary mirror of the platform-level traceability now mandated by the 2026 IT Rules (GAUTAM, 2026).
4. **Independent forensic verification.** In serious criminal cases, verification should be mandatory, court-ordered, and conducted by accredited examiners whose methodology and error rates are disclosed (Shamov, 2024).
5. **AI detector transparency.** Where detection tools are used, they should disclose: methodology; training data and limitations; measured error rates (in-distribution *and* out-of-distribution); validation studies; and confidence levels (Lu & Ebrahimi, 2024; Rawat et al., 2026). A detector that cannot state its own failure rate cannot state anything else — the machine must prove its own provenance.
6. **Judicial AI-evidence protocols.** The Supreme Court, and where appropriate the High Courts, should formulate practice directions establishing the P-R-O-V-E-N-A-N-C-E inquiry, graduated provenance standards, and the preliminary-hearing model for contesting authenticity (Grossman & Grimm, 2025; LaMonaga, 2020).
7. **Right of adversarial technological challenge.** Every party should have the right to challenge the *source, model, methodology, metadata, chain of custody, and expert conclusion* underlying digital evidence — including the right to court-appointed technical assistance where resources are unequal (Shamov, 2024).

XXIV. Conclusion — The Court Must Authenticate the Journey, Not Merely the Destination

The temptation is to end as such papers usually end: with the observation that AI presents challenges and that therefore legislation is necessary. This paper declines that comfort, because it is not enough, and because it is not true. The law has never lacked statutes; it has lacked a conceptual frame adequate to the artefact before it.

The central question of digital evidence can no longer be whether a file is authentic. In an age in which machines can manufacture convincing realities, **authenticity is only the beginning of the inquiry.** The law must ask where the evidence came from, how it came into existence, whether its journey can be reconstructed, whether its transformation can be verified, and whether its meaning can survive adversarial scrutiny. These are provenance questions, and they are the questions the BSA — despite its genuine modernisation — has not yet learned to ask (Goyal & Mali, 2025; Grossman & Grimm, 2025).

The Deepfake Paradox will not be repealed by technology: detection will always lag generation, and the liar's dividend will always be cashed(Schiff et al., 2024). What the law can do is change the currency in which evidence is traded — from appearance to genealogy, from authenticity to verifiability.

The future of digital evidence will therefore belong not to the evidence that looks real, but to the evidence whose provenance can be proved. And so, finally, the closing proposition, which is the whole paper in a sentence: **In the age of generative AI, the law's task is not merely to distinguish the real from the fake. It is to build a system in which the truth remains capable of being proved.** That is the reconstruction this paper has attempted — from authenticity to provenance, from the object to its journey, from the destination to the road.

References:

1. Chesney, R., & Citron, D. K. (2018). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. https://doi.org/10.2139/ssrn.3213954
2. Collomosse, J., & Parsons, A. (2024). To Authenticity, and Beyond! Building Safe and Fair Generative AI Upon the Three Pillars of Provenance. *IEEE Computer Graphics and Applications*, *44*(3), 82–90. https://doi.org/10.1109/mcg.2024.3380168
3. Dhawan, N. (2026). RESEARCH PAPER ON ADMISSIBILITY OF AI-GENERATED EVIDENCE: A COMPARATIVE STUDY OF U.S., EU, AND U.K.

LEGAL FRAMEWORKS. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.18876339

4. Dubey, V. (2017). Admissibility of Electronic Evidence: An Indian Perspective. *Forensic Research & Criminology International Journal*, *4*(2). https://doi.org/10.15406/frcij.2017.04.00109
5. Faqir, R. (2024). THE EXCLUSIONARY RULE OF AI-ENHANCED DIGITAL EVIDENCE IN THE UNITED STATES AND UAE: A COMPARATIVE ANALYSIS. *Journal of Southwest Jiaotong University*, *59*(1). https://doi.org/10.35741/issn.0258-2724.59.1.7
6. Farid, H. (2022). Creating, Using, Misusing, and Detecting Deep Fakes. *Journal of Online Trust and Safety*, *1*(4). https://doi.org/10.54501/jots.v1i4.56
7. Finck, M. (2026). Chapter IV: Transparency Obligations for Providers and Deployers of Certain AI Systems. In *Oxford University Press eBooks* (pp. 464–481). Oxford University Press. https://doi.org/10.1093/law/9780198925705.003.0005
8. Gambín, Á. F., Yazidi, A., Vasilakos, A. V., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: current and future trends. *Artificial Intelligence Review*, *57*(3). https://doi.org/10.1007/s10462-023-10679-x
9. GAUTAM, K. (2026). Deep Fake and the New IT Amendment. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.18987693
10. Goyal, H. R., & Mali, M. P. (2025). Digital Evidence under the Indian Legal System – Study of Problems and Perspective. *Journal of Neonatal Surgery*, *14*, 10180–10189. https://doi.org/10.63682/jns.v14i32s.9809
11. Grossman, M., & Grimm, P. (2025). Judicial Approaches to Acknowledged and Unacknowledged AI-Generated Evidence. *Science and Technology Law Review*, *26*(2). https://doi.org/10.52214/stlr.v26i2.13890
12. Helberger, N., & Diakopoulos, N. (2022). The European AI Act and How It Matters for Research into AI in Media and Journalism. *Digital Journalism*, *11*(9), 1751–1760. https://doi.org/10.1080/21670811.2022.2082505
13. K, S. M., & M.B, S. (2026). New Criminal Laws in India and the Rise of the Digital Court System Leveraging Ai Technology. *Advanced International Journal for Research*, *7*(2). https://doi.org/10.63363/aijfr.2026.v07i02.3810
14. Kandati, B. (2026). From Detection to Provenance: A Deterministic Architecture for Authenticating AI Generated Content. *International Journal of Artificial Intelligence*

Data Science and Machine Learning, *7*, 238–240. https://doi.org/10.63282/3050-9262.ijaidsml-v7i1p139

15. Koutlis, C., & Papadopoulos, S. (2024). DiMoDif: Discourse Modality-information Differentiation for Audio-visual Deepfake Detection and Localization. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2411.10193
16. KUL, M. A. (2026). The Digital Airworthiness Protocol: Establishing a Chain of Custody for Visual Reality. In *Zenodo (CERN European Organization for Nuclear Research)*. European Organization for Nuclear Research. https://doi.org/10.5281/zenodo.18441518
17. LaMonaga, J. P. (2020). A Break From Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes. *eYLS (Yale Law School)*, *69*(6), 5. https://digitalcommons.wcl.american.edu/aulr/vol69/iss6/5
18. Lu, Y., & Ebrahimi, T. (2024). Assessment framework for deepfake detection in real-world situations. *EURASIP Journal on Image and Video Processing*, *2024*(1). https://doi.org/10.1186/s13640-024-00621-8
19. Maras, M., & Alexandrou, A. (2018). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, *23*(3), 255–262. https://doi.org/10.1177/1365712718807226
20. Meding, K., & Sorge, C. (2024). What constitutes a Deep Fake? The blurry line between legitimate processing and manipulation under the EU AI Act. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2412.09961
21. Moruzzi, C., Tallyn, E., Liddell, F., Dixon, B., Collomosse, J., & Elsden, C. (2025). *Content Authenticities: A Discussion on the Values of Provenance Data for Creatives and Their Audiences*. 128–140. https://doi.org/10.1145/3698061.3726918
22. Nitol, A. F. (2026). Navigating the Legal Landscape of Deepfake Technology in Indian Political Dimension. *Journal of Artificial Intelligence Machine Learning and Data Science*, *9*(1), 3313–3322. https://doi.org/10.51219/jaimld/akshat-gupta/666
23. Pawelec, M. (2022). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, *1*(2), 19. https://doi.org/10.1007/s44206-022-00010-6
24. Rawat, R., Dibouliya, A., Chaturvedi, N., Rawat, A., rawat, hitesh, & Dangi, C. S. (2026). A data analytics based federated biometrics for deepfake fraud detection in

financial multimedia systems. *Discover Computing*, *29*(1). https://doi.org/10.1007/s10791-026-10073-5

25. Schiff, K. J., Schiff, D., & Bueno, N. S. (2024). The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability? *American Political Science Review*, *119*(1), 71–90. https://doi.org/10.1017/s0003055423001454
26. Shamov, O. (2024). FROM SYNTHETIC REALITY TO JUDICIAL REALITY: PROCEDURAL CHALLENGES OF "DEEPFAKES" AND THEIR SOLUTIONS IN THE AGE OF ARTIFICIAL INTELLIGENCE. *Entrepreneurship Economy and Law*, *3*, 51–55. https://doi.org/10.32849/2663-5313/2024.3.09
27. Sharma, A. (2020). DEEPFAKES, DIGITAL MANIPULATION AND THE FUTUTRE OF EVIDENTIARY INTEGRITY. In *REDSHINE Publication eBooks*. https://doi.org/10.25215/9358795115.15
28. Singh, B., & Bamb, A. (2021). The Dichotomy of the 65B Certificate. *Christ University Law Journal*, *10*(1), 73–96. https://doi.org/10.12728/culj.18.4
29. Singh, R. (2026). Legal Liability for "Deepfake" Media under Digital India Act. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.18902836
30. Srinivas, M. K. (2026). New Criminal Laws in India and the Rise of the Digital Court System Leveraging AI Technology. *MyPrints@UOM (Mysore University Library)*.
31. Tortora, L. (2024). Beyond Discrimination: Generative AI Applications and Ethical Challenges in Forensic Psychiatry. *Frontiers in Psychiatry*, *15*, 1346059. https://doi.org/10.3389/fpsyt.2024.1346059
32. Twomey, J. G., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLoS ONE*, *18*(10). https://doi.org/10.1371/journal.pone.0291668
33. Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408
34. Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. *arXiv (Cornell University)*, *22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402