

The Label Is Not the Truth

What India and Europe’s new synthetic-media rules can prove, what technology cannot, and why trust needs an architecture

Priyanshu Choudhary | 10 October 2026 | Approximately 7 minutes read

Imagine receiving a video of a company’s chief executive announcing that the business is insolvent. The face is familiar. The voice sounds right. The background is the same office used for every quarterly update.

There is no label on the video. Should you distrust it?

Now change one detail. The video carries a notice saying “synthetically generated”. Should you ignore it?

Neither answer is safe. An unlabelled video may be genuine. A labelled synthetic video may accurately recreate a real statement, translate a speaker with permission, or form part of an obvious satire. The label tells us something important about how the content was made. It does not, by itself, tell us whether the message is true.

That distinction has become legally important. India’s new rules on synthetically generated information took effect in February 2026. The European Union’s transparency duties for certain AI systems began applying in August 2026. Both regimes try to make synthetic content more visible. Their methods differ, but they share an assumption: people make better decisions when the artificial origin of content is disclosed.

The assumption is sensible. It is also incomplete. To see why, we need to separate four ideas that online conversations often collapse into one: origin, integrity, authenticity and truth.

A label answers only one question

The easiest way to understand the problem is to ask four different questions about the same piece of media.

1. **Origin:** Was this image, audio or video captured, generated or materially altered by a machine?
2. **Integrity:** Is the file before me the same file that left the claimed source, or has it changed along the way?
3. **Authenticity:** Is the person, organisation or device making the claim really who it says it is?
4. **Truth:** Does the message correspond to what actually happened?

A visible AI label mainly helps with the first question. A cryptographic provenance record can help with the first three. The fourth question remains stubbornly human and contextual.

A genuine photograph can be old, selectively cropped or paired with a false caption. A correctly signed video can contain a lie spoken by the real person who signed it. Conversely, an AI-generated animation may explain a real scientific process with complete accuracy. The method of production and the truth of the proposition are related, but they are not the same thing.

What India now requires

India's Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 are binding subordinate legislation. [The notified rules](#) came into force on 20 February 2026. MeitY later published an explanatory FAQ, which is useful guidance but expressly states that it does not replace or alter the legal text.

The new Rule 3(3), as explained in the [MeitY FAQ](#), applies to intermediaries that enable or facilitate the creation, modification, publication or dissemination of synthetically generated information. For permitted synthetic audio, visual and audiovisual content, the framework requires a clear and prominent label. It also requires permanent metadata or a technical provenance mechanism, to the extent technically feasible, including a unique identifier capable of identifying the content as synthetic and the computer resource used to create or alter it.

The framework also contains an anti-tampering obligation. An intermediary must not enable the suppression or removal of the required label, metadata or identifier. Significant social media intermediaries face an additional layer: before publication, they must obtain a user declaration about whether the content is synthetic and use reasonable and appropriate technical measures to verify that declaration.

This is not merely a notice-and-takedown model. It tries to build traceability into the content lifecycle: creation, upload, verification, display and response. That is ambitious. It also makes technical feasibility and platform implementation central legal questions.

The European Union separates the machine signal from the human notice

Article 50 of the EU AI Act began applying on 2 August 2026. The [European Commission's guidelines](#) distinguish between duties imposed on providers and duties imposed on deployers.

Providers of generative AI systems must make their outputs detectable as artificially generated or manipulated through effective, reliable, robust and interoperable machine-readable marking. Deployers must clearly disclose deepfakes to people, at the latest upon first exposure. The Commission's FAQ makes the separation explicit: a deployer cannot rely only on an embedded machine-readable mark to satisfy the human-facing disclosure duty.

The legal obligation comes from the AI Act. The Commission guidelines interpret its operation. The [Code of Practice on Transparency of AI-Generated Content](#) is voluntary, although it offers a practical route for signatories to demonstrate compliance. These categories matter. Law, guidance and a voluntary code do not carry the same authority.

The EU approach therefore recognises two audiences. Machines need a signal they can detect and preserve. People need a disclosure they can actually notice and understand. India's approach reaches a similar pairing through prominent labels and provenance mechanisms, while placing distinctive verification duties on significant social media intermediaries.

What provenance actually does

"Provenance" can sound like a sophisticated synonym for a watermark. It is broader than that.

A watermark embeds a signal in the content itself. Metadata travels alongside the content. A provenance system can combine claims about creation and editing history with hashes, digital signatures, timestamps

and information about the signing entity. A verifier can then check whether the claims remain bound to the asset and whether the signature validates under a recognised trust model.

The [C2PA Content Credentials specification](#) is an important example. It provides an architecture for cryptographically verifiable assertions about a media asset's source and history. Its own scope is careful: C2PA does not assign a value judgment that provenance information is "good" or "bad". Validation shows that assertions are associated with the asset, correctly formed and tamper-evident. It does not transform every assertion into truth.

Think of a sealed evidence bag. The seal can show whether the bag was opened after it entered custody. It cannot prove that the item placed inside was described honestly in the first place. Cryptography protects the chain. It does not supply the missing judgment.

Why one signal will never be enough

NIST's [report on digital content transparency](#) examines provenance tracking, watermarking, detection and labelling as complementary techniques. The report also identifies practical weaknesses: metadata can be removed; watermarks can be attacked or degraded; detectors may perform poorly on unfamiliar models or transformed media; and bad-faith actors can attach misleading signals to untrustworthy content.

Ordinary distribution adds more friction. A screenshot may discard the original credential. A messaging platform may recompress a video. A short excerpt may separate content from its context. A visible label can be cropped. A detector can produce a probability that looks authoritative without establishing who created the file or what happened in the world.

This does not make labels or provenance pointless. It tells us how to use them. They should be treated as evidence within a verification process, not as an oracle.

The philosophical problem is trust, not merely fakery

For most of human history, seeing and hearing were shortcuts to believing. Photography and recording complicated that relationship, but generative AI changes its scale. The cost of producing plausible evidence is falling, while the cost of checking it is being transferred to everyone who receives it.

That produces two opposite risks. The first is gullibility: believing a synthetic artefact because it looks convincing. The second is the liar's dividend: dismissing genuine evidence as artificial simply because fabrication is possible.

A society cannot solve either problem by placing an "AI" sticker on selected files. Trust is relational. We trust a statement because of who made it, how it travelled, whether independent evidence supports it, and what accountability follows if it is false. Provenance helps reconstruct that relationship. It does not replace it.

The deeper value of transparency law is therefore not that it tells us what to believe. It creates duties to disclose how an artefact came into being, preserves signals that others can inspect, and identifies actors who can be questioned. It turns trust from an intuition into something closer to an auditable process.

A practical architecture for earned trust

Organisations producing or distributing synthetic media should build controls as a chain, because every isolated control has a predictable failure mode.

1. Before generation: define permitted uses, obtain necessary consent or rights, classify high-risk scenarios and assign an accountable owner.

2. At generation: create the human-facing disclosure, attach machine-readable provenance, and bind relevant claims to the asset with an appropriate signature.
3. At publication: preserve credentials, verify user declarations where required, and ensure the disclosure survives the interface in which people first encounter the content.
4. At verification: combine credential validation with source checks, contextual evidence and carefully evaluated detection tools. Record uncertainty rather than converting probabilities into verdicts.
5. At response: preserve evidence, offer rapid reporting routes for impersonation and intimate-content harms, and maintain a process for correction, removal and escalation.
6. Across the lifecycle: test how signals survive cropping, screenshots, compression, reposting and format conversion. Audit who controls signing keys and who can override labels.

This structure also helps legal and technical teams talk to each other. Lawyers can define the disclosure and accountability outcome. Engineers can test whether the signal remains detectable. Product teams can determine whether a person will see and understand it. Governance connects the three.

The label is the beginning of the inquiry

India and the European Union are trying to make synthetic media legible at the moment when generative systems are making it effortless. That is a necessary regulatory move. It should not be confused with a complete solution to deception.

A label can tell us that a machine participated. A provenance record can tell us which claims travelled with a file and whether they were altered. A signature can connect those claims to a credential. None of these can decide, by themselves, whether the story is true.

The right goal is not a world in which every image comes with an instruction about what to believe. It is a world in which important digital claims arrive with enough structure to be questioned.

Sources and status

[Binding law, India: Information Technology \(Intermediary Guidelines and Digital Media Ethics Code\) Amendment Rules, 2026, G.S.R. 120\(E\).](#)

[Official guidance, India: MeitY FAQs on the 2026 amendments. The FAQ expressly states that it is not a legal document.](#)

[Binding law and official guidance, EU: Article 50 of the AI Act as explained in the Commission transparency guidelines and FAQ.](#)

[Technical report: NIST AI 100-4, Reducing Risks Posed by Synthetic Content.](#)

[Technical standard: C2PA Content Credentials specification.](#)

Editorial note: This article provides general educational analysis, not legal advice. Legal duties depend on role, jurisdiction, scope, exceptions and the facts of a particular deployment.