

DIGITAL INTEGRITY · SYNTHETIC MEDIA · AI EVIDENCE

When Digital Evidence Can Be Manufactured

An evidence-led briefing on synthetic media, synthetic identities, provenance and the standards of verification required before digital material is relied upon.

CARRATU INTERNATIONAL / 9 SEPTEMBER 2026



Digital integrity is now a source-chain problem

Digital material can be persuasive without being authentic. Generative systems can manufacture a voice, image, document, biography or supporting narrative quickly enough that appearance is no longer a dependable proxy for origin.

The UK Fraud Strategy identifies generative AI's ability to create deepfakes that impersonate trusted people and organisations as a significant threat. It also records government work on a framework for detecting deepfake media, including fraudulent documents and synthetic audio. ^[1]

The practical problem is therefore broader than deepfake detection. A manipulated item may sit inside an otherwise credible chain of emails, profiles, corporate records, documents and genuine information.

Integrity question: do not ask only “does this look real?” Ask who created it, where it came from, what independent evidence supports it and what decision depends upon it.

Synthetic identity is broader than a fake face

UK identity-proofing guidance distinguishes an impostor from a synthetic identity and notes that synthetic identities can be wholly fictional or built partly from real information. It recommends checking claimed identities against authoritative counter-fraud sources where risk warrants it. ^[2]

The FCA's 2026 review of AI and retail financial services likewise identifies deepfakes and synthetic identities as technologies capable of exploiting onboarding and decision-making processes. ^[10]

For due diligence and investigation, this distinction matters. A genuine document can support a false narrative; a real person can be presented with invented employment or ownership claims; and a legitimate company can be used as cover for an impersonated contact.

Question	What it tests	Why it matters
Authenticity	Whether an item is what it claims to be	A convincing file can still be fabricated or altered.
Provenance	Origin and identifiable transformations	Source history can be more reliable than appearance.
Corroboration	Independent support for the material claim	A repeated claim is not independent evidence.
Context	Chronology, relationships and surrounding facts	A genuine item can still be misleading out of context.

Detection is evidence, not a verdict

The emergence of dedicated deepfake-detection programmes reflects the seriousness of the problem, but no individual detector should be treated as a complete evidential conclusion. Detection systems can be useful indicators; the surrounding source chain and external facts still require examination.

Where a decision could lead to litigation, payment, reputational action or exclusion of a person, a defensible assessment should preserve the original material, record how it was obtained, compare independent sources and distinguish technical indicators from established facts.

Provenance helps, but it does not prove a claim is true

The C2PA Content Credentials standard is designed to preserve provenance as digital material changes, allowing information about source and transformations to be cryptographically verified.^[4] The European Commission's Article 50 guidance requires machine-readable marking for certain AI-generated or manipulated content and disclosure of deepfakes in defined circumstances. Those transparency obligations apply from 2 August 2026.^[3]

Provenance answers a particular question: what can be established about an asset's origin and declared history? It does not automatically establish that every statement depicted or asserted in the asset is factually correct.

Legal material requires particular discipline

The SRA's August 2026 warning notice makes clear that regulated firms remain responsible for AI-assisted legal material and highlights risks including inaccurate outputs, confidentiality failures and inadequate human oversight.^[5]

The Civil Justice Council's 2026 work on AI in court documents identifies witness statements as an area requiring particular attention because AI may reshape or embellish evidence in ways that are not immediately apparent; its June update emphasises preservation of authenticity, integrity and the witness's personal recollection.^[6]

This does not mean AI involvement automatically invalidates evidence. It means the preparation history, source material and human responsibility around consequential evidence may itself become relevant to reliability.

Practical implication: preserve the original, record provenance, identify the proposition that matters, and corroborate that proposition independently.

Authoritative language is not factual reliability

The ICO distinguishes data-protection accuracy from the statistical accuracy of an AI system and warns that users may wrongly rely on generative AI where the risk and impact of hallucinated or incorrect outputs are not understood and controlled. ^[9]

NIST's Generative AI Profile likewise treats generative AI as a risk-management problem spanning design, development, use and evaluation, rather than something solved by a single test. ^[8]

For investigators and decision-makers, an AI-generated summary can accelerate review, but material assertions still need traceable source support.

Assurance should match the consequence of the decision

The UK government's third-party AI assurance roadmap describes independent assurance as a means of verifying the quality and trustworthiness of AI systems, especially where organisations lack equivalent in-house capability. It estimates the UK AI assurance market at about £1.01 billion GVA in 2024 and says it could exceed £18.8 billion by 2035 if adoption barriers are addressed. ^[7]

That does not mean every organisation needs a full technical model audit. Assurance should be proportionate to use. A low-consequence drafting aid and an AI system influencing onboarding, claims, legal evidence or financial decisions present very different evidential and governance requirements.

Control area	Lower-consequence use	Consequential use
Output checking	Editorial review may be sufficient.	Independent source verification should be built into the process.
Identity	Ordinary account controls.	Authoritative identity and fraud-risk checks may be required.
Evidence	Useful working material.	Preserve originals, provenance, chronology and corroboration.
Governance	Basic acceptable-use controls.	Defined ownership, human oversight, audit trail and escalation.

A practical digital-integrity framework

01 Preserve

Retain the original material, metadata and route by which it was received where possible.

02 Identify the claim

Separate what the material appears to show from the specific fact that matters to the decision.

03 Trace the source

Establish origin, publication history, associated accounts, domains, people and relevant transformations.

04 Corroborate independently

Test the material claim against reliable records and sources that do not depend on the questioned item.

05 Assess inconsistencies

Record conflicts in identity, chronology, ownership, language, metadata or external facts.

06 Report confidence

Distinguish confirmed facts, source assertions, technical indicators and unresolved questions.

The objective is not to prove that AI was involved at any cost. It is to establish whether the information is sufficiently reliable for the decision in front of the client.

Carratu capability

Carratu International can investigate suspected synthetic media, synthetic identities, provenance, information integrity and AI-enabled fraud; assess AI-generated or AI-influenced material in context; and review the governance and evidential controls around consequential AI use.

Where an instruction requires source-code review, adversarial model testing, penetration testing, formal conformity assessment or specialist machine-learning engineering, those elements should be handled by suitably qualified technical specialists. Investigative and governance-focused assurance should not be presented as a substitute for accredited or specialist technical testing.

Sources

Primary and institutional sources used for this briefing. Links are clickable in the PDF edition.

- [1] HM Government. [Fraud Strategy 2026 to 2029: disrupting crime, supporting economic resilience and delivering justice](#). 2026.
- [2] Government Digital Service. [How to prove and verify someone's identity](#). current guidance.
- [3] European Commission. [Guidelines on transparency obligations for providers and deployers of AI systems under Article 50](#). 20 July 2026.
- [4] C2PA. [Content Credentials specification 2.2](#). 2025-2026 specification.
- [5] Solicitors Regulation Authority. [Misuse of AI - Warning notice](#). 17 August 2026.
- [6] Civil Justice Council. [Use of AI in preparing court documents - consultation findings and ongoing work](#). June 2026 update.
- [7] Department for Science, Innovation and Technology. [Trusted third-party AI assurance roadmap](#). 3 September 2025.
- [8] NIST. [Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile \(NIST AI 600-1\)](#). updated 8 April 2026.
- [9] Information Commissioner's Office. [Generative AI: accuracy of training data and model outputs](#). current guidance material.
- [10] Financial Conduct Authority. [Review into the long-term impact of AI on retail financial services \(Mills Review\)](#). 2026.

Scope note

This publication is general research and does not constitute legal, regulatory, technical-forensic or investment advice. Detection tools, provenance systems and assurance methods have different purposes and should be selected according to the material, system and decision being examined.