TimeBinder

A GUIDE FOR LAWYERS

How Lawyers Can Prevent AI Forgeries and Backdating

The complete guide, 2026

> Deepfake calls, forged documents, fake evidence, fabricated citations, false IDs and payment redirection: what to watch for, what to do, and which tools to use.

Updated 11 November 2026
Web version with live links: timebinder.io/guides/ai-forgery-guide-for-lawyers/
General information for lawyers in the US, UK and Europe, Australia, Canada and similar jurisdictions. Not legal advice.

Contents

Executive summary 3

Why AI forgery is now a daily risk for law firms 3

The threat map at a glance 4

The nine kinds of AI forgery, and how to respond to each 5

1. Deepfake audio and video: fake partners, clients and executives 5
2. Forged or altered documents and contracts 6
3. Backdating and metadata manipulation 7
4. Fake evidence: photos, screenshots, chat logs and emails 8
5. Fabricated case citations and hallucinated authorities 9
6. Synthetic identities and ID or KYC fraud 10
7. Invoice fraud, business email compromise and trust-account redirection 11
8. Forged signatures 12
9. AI-generated fake expert reports 13

Core protections that work against every kind of AI forgery 14

Hashing and timestamping: a practical answer to backdating 15

Tools by category (verified links) 17

Tools at a glance table 21

Firm checklist 24

Client checklist 25

Sample engagement clause 26

Sample evidence-intake protocol 27

If you suspect a forgery: a short response plan 27

FAQ 29

Glossary 30

Sources and further reading 30

> **How to use this guide.** Start with the executive summary and the one-page checklist. Use the threat map to jump to the type of forgery in front of you. The sample clause and evidence-intake protocol are templates to adapt with your own regulator's rules in mind. Every outside tool and incident is linked to its source and was checked on 11 November 2026.

Executive summary

Generative AI has made it cheap to clone a voice, fake a video call, rewrite a PDF, invent a screenshot, or produce a convincing but non-existent court decision. Lawyers meet these fakes from three directions: **fraudsters who target the firm** (payment-redirection and impersonation scams), **clients and counterparties who hand over doctored evidence**, and **the firm's own tools**, when an AI assistant invents authorities that end up in a filing.

No single tool solves this. AI-detection software gives a probability, not proof. Metadata and file dates can be edited. A signature image can be copied. What works is **layering**: verify people and instructions out-of-band, capture evidence in a way that preserves its original form, record a fingerprint and an independent time for important files *when you first receive them*, check every authority in a primary source, and tell clients what you will and will not do.

KEY TAKEAWAYS

- **Verify the person, not the message.** Call back on a number you already hold before acting on any payment or account-change instruction, however real the voice or video looks.
- **Treat all media as unverified until its origin is known.** Ask for the original file from the original device, in native format, with metadata intact.
- **Fingerprint and timestamp at intake.** A SHA-256 hash plus an independent timestamp shows an exact file existed no later than a recorded time and still matches. It is the practical answer to backdating allegations.
- **Do not rely on detectors alone.** Use them as one signal, record the result, and escalate high-stakes items to a forensic expert.
- **Cite-check everything an AI tool gives you** against a primary source before it reaches a court, a client or an opposing party.
- **Put it in writing.** A short engagement-letter clause and a client checklist prevent most payment-redirection losses.
- **Have a response plan** for a suspected forgery: stop, preserve, escalate, and consider your duties to the client and the tribunal.

This guide is written for lawyers practising in the United States, the United Kingdom and Europe, Australia, Canada and similar jurisdictions. It stays at the level of practice and process. Professional rules, evidence rules and privilege differ, so check your own regulator and court rules before you adopt any template here. It is general information, not legal advice.

Why AI forgery is now a daily risk for law firms

Law firms hold exactly what fraudsters want: large sums in trust or client accounts, urgent deadlines, confidential deals and a culture of acting on instructions quickly. At the same time, lawyers depend on documents, recordings, messages and authorities whose authenticity they have rarely needed to test.

Three changes matter. First, **cloned voices and faces are now cheap**: a short sample of public audio or video can be enough to build a convincing impersonation. Second, **text-and-image generators produce plausible documents on demand**, including letters, invoices, ID photos, screenshots and chat logs. Third, **generative AI research tools can invent legal authorities** that look real. The result is a wider set of ways for something false to enter a file.

Courts have started to respond. In 2023 a New York federal court sanctioned lawyers for filing non-existent cases generated by an AI chatbot. In 2025 a California judge dismissed a civil case after finding that video evidence was AI-generated, and the senior judiciary of England and Wales warned that placing fake citations before a court could have serious consequences for the lawyers involved. These are real, reported decisions, set out with sources below.

A NOTE ON WHAT IS AND IS NOT CLAIMED HERE

Every incident in this guide is taken from a court order, regulator statement or major news report, and each is linked so you can read the source. Where we describe a tool, we describe what its own site says it does. Detection tools, content credentials and timestamps each answer a *different* question; none of them proves a document is true.

The threat map at a glance

Use this table to find the section that matches the problem in front of you. Each type has its own section below with a real incident, red flags, protections and tools.

Nine kinds of AI forgery lawyers meet

Type	What it looks like	Typical exposure for a lawyer	First line of defence
1. Deepfake audio and video	A cloned "partner", client or CEO on a call or voice note asking for a payment or confidential data	Trust-account transfers, wrongful disclosure, impersonated clients	Call-back on a known number; a second approver for payments
2. Forged or altered documents	A contract, will, deed, letter or bank statement edited or generated by AI	Relying on a false document; disputes over which version was signed	Fingerprint the version received; compare against a trusted copy
3. Backdating and metadata manipulation	Files with edited creation dates, back-filled documents, "I wrote this in 2022"	Priority, notice and limitation disputes; credibility of your client	Independent timestamp at intake; preserve native files and metadata
4. Fake photos, screenshots, chats, emails	AI-edited images, fabricated messages, doctored email headers	Tendering false evidence; losing the case; sanctions	Original-device capture, chain of custody, forensic review when stakes are high
5. Fabricated citations	Cases, quotes and statutes that do not exist, produced by an AI tool	Sanctions, costs orders, regulatory referral, reputational harm	Check every authority in a primary source before it leaves the firm
6. Synthetic identities and ID fraud	AI-generated passports, licences and face images used to pass checks	Acting for a fraudster; breach of client-identification obligations	Layered ID checks, liveness and source-of-funds questions

Type	What it looks like	Typical exposure for a lawyer	First line of defence
7. Invoice fraud and payment redirection	Convincing emails and invoices changing bank details; spoofed counterparty accounts	Money lost from trust or client accounts; negligence claims	Verbal verification of every change in payment details
8. Forged signatures	Copied or generated signature images on documents or e-sign requests	Relying on a document the signer never signed	Audit-trail e-signature; witnessing and ID steps for high-value documents
9. Fake expert reports	Reports with invented sources or credentials, or unattributed AI-written analysis	Tendering unreliable evidence; credibility loss	Source-check every reference; confirm credentials; ask experts how AI was used

The nine kinds of AI forgery, and how to respond to each

Each subsection follows the same pattern: what the risk is, a real incident, the red flags, what to do and which tools help. The sections after this one cover the cross-cutting protections that apply to all nine.

1. Deepfake audio and video: fake partners, clients and executives

The risk. A fraudster clones a voice or builds a synthetic face, then uses it in a phone call, voice note or live video meeting to instruct a payment, ask for confidential information or impersonate a client who is "unable to attend in person". The pretext is usually urgency and secrecy. For law firms the most common version targets the trust or client account, or tries to open a matter with a stolen identity.

> **REAL INCIDENT: ARUP, HONG KONG (2024)**
>
> In early 2024 an employee in the Hong Kong office of the engineering firm Arup was invited to a video conference that appeared to include the company's UK chief financial officer and other colleagues. According to Hong Kong police and Arup, all of the other participants were digital fakes. Following the instructions given, the employee made 15 transfers totalling about HK$200 million (about US$25 million) to five bank accounts. The fraud came to light when the employee later checked with the head office. Arup confirmed that "fake voices and images were used". Sources: CNN, 16 May 2024 and The Guardian, 17 May 2024. An earlier voice-only case: in 2019 a UK energy company's chief executive transferred about €220,000 after a caller using AI voice imitation posed as the chief executive of the German parent company (Wall Street Journal, 2019; Forbes, 2019).

Red flags.

- An urgent, confidential payment or change of details, especially one that bypasses your usual approvals.
- A caller or meeting host who refuses to be called back on a known number, or whose video is oddly smooth, repetitive or out of sync with speech.

- A new contact method (a new mobile number, personal email or messaging app) used by someone you normally reach another way.
- Instructions to keep the matter from colleagues, or to skip the second approver.
- A client who "cannot" turn the camera on, show ID live, or take a call on a number you hold.

What to do.

- **Out-of-band call-back.** Confirm any payment, account change or identity step by calling a number from your own records, not one supplied in the message or call.
- **Two-person rule.** Require a second authorised person to approve any transfer above a set threshold or to a new payee, and make that person verify independently.
- **Challenge phrases.** For regular institutional clients, agree a verbal passphrase or a one-time code sent through a separate channel.
- **Treat live video as unverified.** A familiar face on screen is no longer proof of identity. Pair video with a call-back or in-person check.
- **Preserve the call.** Keep call logs, recordings (where lawful), meeting links and invitations. Hash and timestamp any recording you intend to rely on.

Tools. Detection services such as Reality Defender, Pindrop, Resemble AI Detect, Sensity AI and Hive (now branded Deploy) may add a signal on suspect audio or video, as can Amped Authenticate in expert hands. They give probabilities, can be fooled and can produce false positives, so a call-back is still the control that works. Intel's FakeCatcher and TrueMedia.org are worth knowing about but are not general-purpose services for most firms.

2. Forged or altered documents and contracts

The risk. A generative tool can produce a convincing contract, letter, pay slip, bank statement or medical note in seconds, and a PDF editor can quietly change a single clause, amount or date in a genuine document. The danger for a lawyer is relying on, serving or filing a document that is not what it purports to be, or being unable to show which version was actually signed.

Red flags.

- A document that exists only as a screenshot, print-to-PDF or photograph, with no original file.
- Inconsistent fonts, spacing or alignment, odd rounding, or numbers that do not add up.
- Metadata that does not match the story: a "2021" letter created last week, an author field that is blank or generic, or a PDF producer that is an AI or image-editing tool.
- Different versions circulating, with each party insisting theirs is the signed one.
- Refusal to provide the native file or to let you contact the issuer (a bank, an employer, a registry).

What to do.

- **Get the source.** Ask for the native file from the system that produced it, or confirm the document directly with the issuer using contact details you found yourself.
- **Fingerprint on receipt.** Record the SHA-256 hash of every important document as received, then compare later copies against it. See the free file comparison checker for a browser-based way to do this.

- **Version control at signing.** Agree a single execution version, hash it, and circulate only the hash-matched file. Keep the signed original read-only.
- **Compare, do not eyeball.** Use a document-comparison tool between the version you drafted and the version returned for signature.
- **Record how the file reached you**: sender, channel, time and any intermediate steps.

Tools. ExifTool shows embedded metadata, and a hash comparison shows whether two files are byte-for-byte identical. E-signature platforms such as DocuSign and Adobe Acrobat Sign create audit trails for documents signed through them. For certainty about a disputed file, instruct a forensic examiner (for example using software like Amped Authenticate or Magnet Forensics).

---

3. Backdating and metadata manipulation

The risk. Backdating means creating or altering a document so that it appears to date from an earlier time. It has always happened, but AI makes it quicker to produce a plausible "old" document, and file-date editing tools are freely available. In disputes over priority, notice, consent, knowledge, prior use, deadlines or limitation, the question "when did this file really exist?" can decide the case.

A file's creation date, modified date and embedded metadata are **self-reported by the computer that made the file**. Anyone can change them. Email headers can be forged. Screenshots of timestamps can be edited. None of that is evidence of when the content was created unless it is corroborated by something the file's author does not control.

> **WHAT INDEPENDENT TIME EVIDENCE LOOKS LIKE**
>
> - **A trusted timestamp** issued by a third-party time stamp authority (the RFC 3161 model) binds a hash of the file to a time signed by that authority.
> - **A blockchain-anchored timestamp** records a hash in a public ledger whose history nobody can quietly rewrite.
> - **Third-party records**: a filing receipt, a registered post receipt, an upload log on a platform the author does not control.
>
> Each of these shows that *an exact file existed no later than a recorded time*. They do not show who wrote it, or that it is true.

Red flags.

- A document whose internal dates do not match its creation or modification metadata, or whose metadata has obviously been scrubbed.
- A set of "contemporaneous" notes or emails that all share one creation time, or were created after the dispute began.
- A file created in a recent version of software but dated well before that version existed.
- A story that depends on a single document that surfaced only after the other side raised the point.
- A client who cannot produce the original device, account or backup.

What to do.

- **Timestamp at creation and at intake.** Ask clients to fingerprint and timestamp key records (design files, drafts, decision notes, photographs) when they create them. At intake, hash and timestamp the files they give you so that, from that day forward, you can show the file has not changed.
- **Preserve natively.** Collect files by forensic copy or native export, not by opening, forwarding or re-saving. Keep a log of who handled them.
- **Corroborate dates** with sources the author does not control: platform logs, cloud revision history, third-party emails, payment records.
- **Never "tidy up" a file** before preserving it. Opening and re-saving can change metadata.
- **Ask early.** If the other side's evidence depends on a date, ask for native files and metadata promptly and consider a preservation letter.

Tools. OpenTimestamps, FreeTSA, DigiStamp, OriginStamp and TimeBinder all create independent time evidence of different kinds, compared in the timestamping section. ExifTool reads metadata, and Magnet Forensics and similar forensic tools acquire device and cloud data in a way that preserves it.

4. Fake evidence: photos, screenshots, chat logs and emails

The risk. Image generators and editors can change faces, objects and text in a photograph, add or remove injuries, damage or documents, or produce an entire scene. Chat logs and emails can be fabricated in a few minutes, and screenshots are the easiest of all to forge. The worst outcome is not only a lost case: a lawyer who tenders false evidence knowingly or carelessly may face sanctions and professional consequences.

> **REAL INCIDENT: MENDONES V. CUSHMAN & WAKEFIELD (CALIFORNIA, 2025)**
>
> In September 2025 an Alameda County Superior Court judge dismissed a civil case with terminating sanctions after finding that exhibits the plaintiffs filed in support of a summary judgment motion were created or altered with generative AI. Video exhibits purporting to show a witness lacked natural facial expressions, looped, and had mouth movements that did not match the words; the judge also reviewed the exhibits' metadata. Reconsideration was refused in November 2025. Sources: the court's order and NBC News.

Red flags.

- Only a screenshot or an exported image of a conversation, with no way to see the underlying account, thread or device.
- Photos or video that are unusually clean, or have inconsistent shadows, text, hands, reflections or edge artefacts; faces with little expression; looping backgrounds.
- No original file, or a file stripped of camera metadata.
- A sequence of messages with unnatural gaps, identical formatting, or timestamps that fall outside the platform's real behaviour.
- Resistance to a forensic extraction from the device or account.

What to do.

- **Collect at source.** Where it matters, export chats and emails from the account or device with a forensic tool, or have an independent examiner collect them, rather than accepting screenshots.
- **Original device, original file.** For photos and video ask for the original camera file and the device; note device, app and settings.
- **Chain of custody.** Log every person who held the evidence, when, and where it was stored; re-hash whenever it moves.
- **Check provenance data.** Look for Content Credentials and compare metadata, but treat their absence as neutral, not as a finding of fakery.
- **Know the rules.** Evidence rules often allow authentication by hash value (for example Rule 902(14) in the US federal rules: text), but you need a procedure that creates and records that hash at the right time.
- **Escalate for high stakes.** A forensic examiner can assess a disputed image or video and can give evidence on method and reliability.

Tools. FotoForensics, Forensically, InVID-WeVerify verification plugin and TinEye help screen images and video; Content Credentials Verify reads Content Credentials where present; Truepic captures verified media inside its own workflow; Amped Authenticate and Magnet Forensics are specialist forensic tools. For a plain-English companion, see how to prove a photo is not AI and what to do when the other side says your evidence is a deepfake.

5. Fabricated case citations and hallucinated authorities

The risk. General-purpose chatbots and some legal tools generate fluent text that includes cases, quotations or statutory provisions that do not exist, or that exist but do not say what the text claims. The harm is to the client, the court and the lawyer's own standing. Cases have now been reported in every major common-law jurisdiction.

REAL INCIDENTS: UNITED STATES, ENGLAND, CANADA AND AUSTRALIA

- **United States: *Mata v. Avianca, Inc.* (S.D.N.Y., 22 June 2023).** Lawyers filed a brief citing non-existent judicial opinions with fake quotes generated by ChatGPT, and continued to stand by them after the court questioned them. The court imposed a joint US$5,000 penalty and ordered letters to the judges falsely named as authors. Opinion and order on sanctions.
- **England and Wales: *Ayinde v London Borough of Haringey; Al-Haroun v Qatar National Bank* [2025] EWHC 1383 (Admin) (6 June 2025).** A Divisional Court dealt together with two matters where false citations or quotations had been put before the court, and warned that those responsible for training, supervising and regulating lawyers must ensure professional duties are understood. Judgment.
- **Canada: *Zhang v Chen*, 2024 BCSC 285.** A British Columbia lawyer included two non-existent cases generated by ChatGPT in a notice of application. The court declined special costs because there was no intent to deceive, but ordered her to personally bear the costs of the extra effort caused. Decision on CanLII.
- **Australia: *Dayal* [2024] FedCFamC2F 1166.** A Victorian solicitor tendered a list and summary of authorities generated by AI that contained inaccurate citations. The Victorian Legal Services Board later varied his practising certificate so that he could practise only as a supervised employee solicitor, with no trust-money handling. Regulator statement; The Guardian.

Red flags.

- A case name and citation you cannot find in a primary database.
- A quotation that does not appear in the cited judgment, or a judge, court or year that does not fit.
- Neat, confident summaries with no pinpoint reference.
- A junior, a contractor or an opposing party who cannot say where an authority came from.

What to do.

- **Check every authority in a primary source** before it leaves the firm: read the case, confirm the quote, run it through a citator, and confirm it is still good law.
- **Adopt an AI-use policy** that states which tools are approved, what may be entered into them (confidentiality), and that a named lawyer is responsible for verifying every output.
- **Keep a verification log** for each filing: who checked each authority, and when.
- **Supervise.** Treat AI-generated drafts like work by a very fast, unreliable junior: review closely.
- **If an error slips through,** correct it promptly and candidly with the court and the client. Courts have treated early candour more favourably than doubling down.

Tools. Citators and research platforms: Westlaw (KeyCite citator), Lexis+ (Shepard's citations), CoCounsel (Thomson Reuters), vLex (now part of Clio). Cite-checking inside Word: Clearbrief. Free primary-law databases to confirm that a case exists: CourtListener citation lookup (US), BAILII (UK and Ireland), CanLII (Canada), AustLII (Australia and the region).

6. Synthetic identities and ID or KYC fraud

The risk. AI can generate realistic passports, driving licences, utility bills and selfie photos, and can produce deepfake faces that pass naive liveness checks. A law firm may be instructed by someone who is not who they claim to be, usually to move or hide money, transfer property or obtain documents. Firms must meet their own client-identification and anti-money-laundering obligations; failing them is both a risk and a regulatory issue.

> **REAL INCIDENT: ONLYFAKE (2024)**
>
> In February 2024 the news outlet 404 Media reported that an underground service called OnlyFake was using AI to generate realistic photographs of fake IDs for about US$15, and that its reporters used a generated ID to pass an identity check at a cryptocurrency exchange. The operator later pleaded guilty in a US federal court to selling more than 10,000 fake identity-document images, primarily to get around know-your-customer checks at banks and crypto exchanges. Sources: 404 Media and BleepingComputer.

Red flags.

- A client you have never met who wants speed, secrecy and a large transfer, with no clear business reason.
- ID photographs taken "on a carpet" or table with perfect lighting, inconsistent fonts or microprint, or a face that does not match a live video check.
- An address history, employer or phone number that cannot be corroborated.

- A refusal of in-person or notarised verification.
- Source-of-funds explanations that change when questioned.

What to do.

- **Layer your checks.** Combine document verification, biometric or liveness checks, database checks and independent corroboration. Avoid relying on one uploaded image.
- **Use an electronic ID or in-person step** when stakes are high, and follow your regulator's identification rules.
- **Ask source-of-funds and source-of-wealth questions** and document the answers.
- **Re-verify on change.** Re-check identity when a client's instructions, contact details or beneficiary change.
- **Report suspicions** as required by your jurisdiction's anti-money-laundering rules.

Tools. Identity verification services include Entrust (formerly Onfido), Persona, Jumio and Veriff. For legal and property practices in Australia, InfoTrack and First AML provide AML and identity workflows. Each vendor will describe its own liveness and document checks; none is infallible.

7. Invoice fraud, business email compromise and trust-account redirection

The risk. The classic scam is an email, sometimes from a genuinely compromised account, saying that bank details have changed. Today the email can be flawlessly written in the right tone, attach a convincing invoice, and be followed by a cloned-voice call to "confirm". Settlement funds, deposits and trust-account transfers are prime targets because the amounts are large and the timing is predictable.

> **REAL DATA: FBI INTERNET CRIME REPORT**
>
> The FBI's Internet Crime Complaint Center reported business email compromise losses of US$3,046,598,558 in its 2025 report (published in 2026), up from US$2,770,151,146 in 2024, making it one of the largest cyber-enabled fraud loss categories. Source: FBI IC3 2025 Internet Crime Report. The IC3 is also where US victims report fraud: ic3.gov.

Red flags.

- Any change to bank details, however small, particularly close to settlement or payment date.
- A sender address that differs by one character from the real one, or a reply-to that does not match.
- Pressure to pay today, outside normal channels, or to a different bank or country.
- An invoice that arrives from the right person but at an unusual time, or with a PDF whose metadata shows it was created seconds earlier.
- A request that you not call to verify.

What to do.

- **Verbally verify every change in payment details** by calling a number you already hold (not one in the email). Make this a firm rule with no exceptions for urgency.

- **Put your own payment details in the engagement letter** and state that you will never change them by email alone (see the sample engagement-letter clause).
- **Dual authorisation** for new payees and for transfers above a threshold.
- **Warn clients early** that your firm's bank details will not change by email, and provide a way to check.
- **Secure the mailbox.** Multi-factor authentication, mailbox-rule monitoring and a clear process for reporting a suspected compromise.
- **If it happens,** contact your bank immediately to try to recall the funds, preserve the evidence, and notify the relevant authorities and insurer promptly.

Tools. There is no single tool. Technical controls (multi-factor authentication, email authentication such as SPF, DKIM and DMARC, domain monitoring) and human procedure do the work. DocuSign and Adobe Acrobat Sign reduce the risk of altered contracts, but the payment instruction still needs a call-back. Local reporting channels include the FBI IC3 (US) and Scamwatch (Australia).

8. Forged signatures

The risk. A signature image is easy to copy from any signed document, and generative tools can imitate handwriting. In an e-signature environment the risk shifts to who controlled the email account or device that completed the signing. In either case a lawyer may rely on a document the apparent signer never signed.

No specific reported incident is cited here because the underlying risk is long-standing and the reported AI-specific cases are still emerging; check your own jurisdiction's case law.

Red flags.

- A signature that is pixel-identical across several documents (a pasted image).
- A signing certificate or audit trail that shows an unfamiliar IP address, device or email domain.
- A counterparty who signs by email attachment, scanned image or "wet signature" photo, with no witness.
- Signing activity at unusual hours or in very rapid succession.

What to do.

- **Prefer audit-trail e-signing** with identity checks proportionate to the document's value, and keep the completion certificate with the signed file.
- **Check the audit trail** against what you know: who was sent the envelope, from where, and how they were verified.
- **Use witnessing, notarisation or in-person execution** where the law or the stakes call for it.
- **Fingerprint the final executed version** so you can later show exactly which file was signed.
- **For high-risk matters,** confirm by call-back that the signer actually signed.

Tools. DocuSign and Adobe Acrobat Sign generate audit reports. A timestamp does not replace a signature: it shows when a file existed, not who signed it (see timestamp vs electronic signature).

9. AI-generated fake expert reports

The risk. Expert reports, valuations, medical summaries and technical analyses can include invented sources, fabricated credentials or unattributed AI-generated text. A false report can mislead the tribunal and damage the expert's credibility along with your case.

> **REAL INCIDENT: KOHLS V. ELLISON (MINNESOTA, 2025)**
>
> In January 2025 a federal judge in Minnesota excluded an expert declaration filed in a case about a deepfake law, after it emerged that the declaration, drafted with help from an AI model, contained citations to academic sources that did not exist. The court said the errors undermined the declaration's credibility and refused permission to file an amended version. Sources: the court's order and Reuters.

Red flags.

- References you cannot locate, or whose titles and authors look slightly wrong.
- Credentials or memberships you cannot confirm with the issuing body.
- Uniform, generic prose with no reasoning specific to the facts.
- An expert who is reluctant to disclose their method, tools or instructions.

What to do.

- **Check every cited source** and every credential before the report is served.
- **Ask experts in writing** whether any AI tool was used in preparing the report and how its output was verified.
- **Preserve the instructions and drafts** exchanged with the expert, hashed and timestamped so the record of what was supplied and when cannot be disputed.
- **Compare to the expert's earlier work** for style and reasoning consistency.
- See also timestamping expert reports.

Core protections that work against every kind of AI forgery

The nine types look different, but the defences repeat. Build these into the firm's routine and most of the risk falls away.

Verify people and instructions out-of-band

Treat every inbound instruction as unverified until you confirm it through a channel the sender does not control. A call-back to a number already on file defeats cloned voices, fake video meetings and spoofed email. Make it routine, written into your procedures and applied to partners as well as clients, so nobody is embarrassed by following it.

- Confirm payment instructions and any change to payment details by phone on a known number.
- Agree a verification method with regular clients at the outset (passphrase, code through a separate channel).
- Require two approvers for new payees and for large transfers.
- Never accept "I'm travelling, just trust me" as a reason to skip a step.

Chain of custody and evidence handling

A chain of custody is a written record of who held a piece of evidence, when, where it was stored and what was done to it. Without it, even genuine evidence can be attacked. Record the source and date of receipt, keep an untouched original, work on copies, and re-hash when a file moves.

Preserve metadata and native files

Ask for native files, not screenshots or re-saved PDFs. Do not forward, open and re-save, or "clean up" a file before it has been preserved. Export metadata with a tool such as ExifTool and store the output with the file. Remember that metadata is useful but self-reported and can be edited.

Hash and timestamp when you receive the file

A hash is a short fingerprint of a file's exact contents: change one character and the hash changes completely. A timestamp from an independent source ties that hash to a time. Together they let you show later that a file is unchanged and existed no later than a recorded time. This is the most direct defence against backdating and "that's not the version I signed" disputes. The next section compares the options.

Use provenance data where it exists (C2PA and Content Credentials)

Content Credentials, built on the open C2PA (Coalition for Content Provenance and Authenticity) standard, can record how a photo, video or image was captured and edited, signed so tampering is detectable. You can inspect credentials on a file with Content Credentials Verify. Support is growing in cameras, phones, editing software and AI generators, but credentials are only present if the tool added them and are often stripped when a file is uploaded or re-saved. Their absence proves nothing; their presence and validity is a helpful signal about origin.

E-signature audit trails

E-signature services such as DocuSign and Adobe Acrobat Sign keep a record of who was sent an envelope, when it was opened and signed, and from which device or address. Keep the completion certificate with the executed document and compare it against what you know about the signer.

Engagement-letter clauses and client education

Most losses happen because a client was not told what to expect. State in the engagement letter how you will and will not communicate payment details, ask clients to preserve and provide evidence in native form, and explain how to verify a message from you. A one-page client checklist (below) helps. A sample clause is provided as a template to adapt.

Use detection tools carefully

Deepfake and AI-content detectors are probabilistic. They can miss new generators and can flag genuine material. Use them as one signal, record what you ran and the result, never treat a "not detected" result as authentication, and engage a qualified forensic examiner for anything high-stakes.

Competence, candour and confidentiality, in general terms

Professional rules differ by country and by state or province, but most turn on the same ideas, and recent cases reflect them:

- **Competence.** Understand the tools you use and their limits, including the risk of fabricated output. ABA Formal Opinion 512 (US, 2024) and the Law Society of England and Wales' guidance discuss this.
- **Candour to the tribunal.** Do not put material before a court that you know, or ought to know, is false, and correct errors promptly. The cases above (*Mata*, *Ayinde*, *Zhang v Chen*, *Dayal*) all concern this duty.
- **Confidentiality.** Check what a tool does with the data you enter before using it with client material. Hashing and timestamping approaches that keep the file on your device can help, but check your own obligations.
- **Supervision.** Responsibility for AI-assisted work stays with the lawyer; make sure staff are trained and policies are written.
- **Client communication.** Tell clients about material risks and about how you will verify instructions.

Check your own regulator's rules and any court practice directions on AI and evidence. This guide does not state what the law requires in any particular place.

Hashing and timestamping: a practical answer to backdating

Hashing and timestamping do one specific job: they let you show that a file you hold today is **exactly the file that existed no later than a recorded time**. They do not show who created the file, that its content is true, or whether AI was used to make it. That limit matters, and it is why this is one layer among several.

How it works in four steps

1. **Hash.** Software computes a SHA-256 fingerprint of the file. Anyone can recompute it later; if even one bit has changed, the fingerprint is different.

2. **Time-stamp the hash.** A third party signs the hash with a time (an RFC 3161 authority) or the hash is anchored in a public blockchain.
3. **Keep the proof.** Store the proof file or certificate next to the original.
4. **Verify later.** Re-hash the file, compare with the recorded hash, and verify the timestamp.

Timestamping options compared

Option	How it records time	Access and cost	Things to consider
FreeTSA	RFC 3161 timestamp token signed by a public authority	Free	No stated service guarantees; consider a commercial TSA for important matters
DigiStamp	RFC 3161 timestamp from a commercial authority, with archive	Paid; see pricing page	Established commercial TSA; check validation path for your jurisdiction
OpenTimestamps	Hash anchored in the Bitcoin blockchain with an open proof file	Free, open source	Technical to use; proof is independently verifiable by anyone with the tools
OriginStamp	Blockchain timestamping through a web service and API	Commercial	Check current product scope; now emphasises an enterprise e-archive
TimeBinder	Browser makes the SHA-256 fingerprint; the fingerprint is anchored to the Bitcoin blockchain; downloadable certificate	First Pooled TimeBind free, no card; then from US$5; ASAP is US$15	The file stays on your device. Pooled records are batched and checked through TimeBinder; ASAP uses its own Bitcoin transaction

Where TimeBinder fits, honestly

TimeBinder is one way to do the hash-and-timestamp step without installing software. Your browser creates the SHA-256 fingerprint, the file stays on your device, and TimeBinder records the fingerprint and time and issues a certificate you can keep with the file. It shows that an exact file existed no later than a recorded time and that a copy shown later still matches. It does not show who made a file, who owns it, that its contents are true, or whether AI was used. The first Pooled TimeBind is free with no card, then from US$5; the free file comparison checker lets you compare two files in your browser. How the courts treat any timestamp depends on the jurisdiction and the case, so treat it as one piece of evidence and take advice. See the legal timestamping hub for guides written for law firms.

For firms. Some firms offer TimeBinds to clients and resell them at the firm's own margin, or refer clients for a 20% referral commission. Neither is necessary to use anything in this guide. If you are considering it, see offering timestamping to clients and the professional-conduct checklist for reselling.

Related plain-English answers: how to prove a contract wasn't edited after signing; how to tell if a file has been changed; hash timestamp vs notarisation; how to prove a file existed before a date.

Tools by category (verified links)

Every link below was fetched on 11 November 2026. Where a site returned a bot-check or "access denied" page to our automated fetch, it is marked as a real block page rather than a failure, and the URL is a genuine vendor address. Descriptions are neutral summaries of what each vendor says it does, not endorsements, and tools change: confirm current features and pricing with the vendor. Links are plain links; no affiliate links are used.

Deepfake and synthetic media detection

Tool	What it does	Keep in mind
Reality Defender realitydefender.com Commercial (contact sales)	Detection platform for synthetic audio, video and images, with products aimed at calls and meetings as well as uploaded files.	Returns a probability-style assessment, not proof. Aimed at organisations rather than individual lawyers.
Hive (now branded Deploy) hivemoderation.com Commercial	Classifiers that flag AI-generated and deepfake images, video and audio, mainly used by trust-and-safety and enterprise teams.	Probabilistic output; false positives and negatives occur. Vendor has rebranded, so check current product names.
Sensity AI sensity.ai Commercial (trial offered)	Forensic-style deepfake detection for video, images and audio through a web app and API, with analyst reports.	Treat reports as one input to a wider authenticity assessment; how a court weighs any such report is a matter for the court and your expert.
Intel FakeCatcher intel.com Research / enterprise	Intel research project for real-time detection of synthetic face video, described on Intel's research pages.	Not a public upload service; reported as an enterprise-oriented detector. Listed for awareness, not as a tool most firms can use directly.
TrueMedia.org truemedia.org Non-profit initiative (request access)	Non-partisan, open-source initiative whose platform helps people assess whether images, audio and video are manipulated.	Access is by request and the project has changed form, so check availability before relying on it.
Pindrop pindrop.com Commercial (enterprise)	Voice security and deepfake detection for calls, virtual meetings and contact centres.	Built for large call volumes; most useful to firms with a call-centre or high-risk voice workflow.
Resemble AI Detect resemble.ai Commercial	Detection of synthetic audio, video and images, offered by an AI voice company, with an API and real-time options.	Probabilistic; best used alongside human verification such as call-backs.
Google SynthID deepmind.google Google DeepMind (watermarking research)	Google's watermarking approach for content made with its own generative tools, with a related detector.	Only identifies content carrying Google's SynthID watermark. A clean result says nothing about content from other tools.

Content provenance

Tool	What it does	Keep in mind
C2PA (Coalition for Content Provenance and Authenticity) c2pa.org Open standard	Standards body behind the C2PA specification for Content Credentials: signed metadata recording how a piece of media was made and edited.	Credentials only exist if the camera, software or platform adds them, and they can be stripped when files are re-saved or uploaded.
Content Credentials Verify contentcredentials.org Free	Free web tool that reads Content Credentials from a file and shows the recorded provenance, if any.	No credentials found does not mean a file is fake; it usually means none were attached or they were removed.
Content Authenticity Initiative contentauthenticity.org Free information / community	Adobe-founded community promoting Content Credentials, with explainers and open-source tools.	Educational; adoption by cameras and platforms is still uneven.
Adobe Content Authenticity contentauthenticity.adobe.com Adobe service	Adobe's service for attaching and inspecting Content Credentials on creative work.	Useful for new work created or edited inside a credential-aware workflow, not for legacy files.
Truepic truepic.com Commercial	Photo and video authentication at capture, mainly for inspections, insurance and similar workflows.	Protects media captured inside its workflow; it does not retroactively authenticate files made elsewhere.

Timestamping and hashing

Tool	What it does	Keep in mind
TimeBinder timebinder.io First Pooled TimeBind free, no card; then from US$5	Browser-based service that records a file's SHA-256 fingerprint with a time and anchors it to the Bitcoin blockchain. The file stays on your device; a certificate is downloadable.	Shows an exact file existed no later than a recorded time and still matches. It does not show who made it, who owns it, that it is true, or whether AI was used.
OpenTimestamps opentimestamps.org Free, open source	Free, open-source protocol and tools that anchor file hashes to the Bitcoin blockchain; proofs can be checked without trusting a company.	Command-line and developer oriented; upgrading and verifying a proof needs some technical comfort.
RFC 3161 (Time-Stamp Protocol) rfc-editor.org Open standard	The IETF standard for trusted timestamp authorities (TSAs) that sign a hash together with a time.	A standard, not a service; evidential weight depends on the authority chosen and how its certificates are kept.
FreeTSA freetsa.org Free	Free RFC 3161 time stamp authority operated as a public service, with web form and command-line options.	Free service with no stated service guarantees; check its terms and consider a commercial authority for high-stakes matters.
DigiStamp digistamp.com Commercial	Commercial RFC 3161 timestamp authority with an archive, pricing page and add-ins for common document software.	Paid; lawyers should check which timestamp format and validation path suits their jurisdiction.
OriginStamp originstamp.com Commercial	Provider of blockchain timestamping (web service and API) and an e-archive product called OriginVault.	Products are shifting toward archiving for enterprises; confirm current offerings.

Forensics and metadata

Tool	What it does	Keep in mind
ExifTool exiftool.org Free, open source	Phil Harvey's command-line tool for reading and writing metadata in images, video, audio and many documents.	Shows what metadata a file claims, which can be edited; it does not prove the metadata is genuine.
FotoForensics fotoforensics.com Free web tool	Web tool that runs error-level analysis and shows metadata to help look at possible edits to an image.	Results need expert interpretation; ELA alone can mislead. Uploading an image shares it with the service, so avoid privileged files.
Forensically 29a.ch Free web tool	Browser-based photo forensics toolkit with clone detection, error-level analysis and metadata viewing.	Runs in the browser; still a screening aid, not an expert opinion.
InVID-WeVerify verification plugin invid-project.eu Free browser extension	Browser plugin built for journalists to verify images and video: keyframes, reverse search and metadata.	Some analysis calls external services, so check what leaves your device before using it on confidential material.
Amped Authenticate ampedsoftware.com Commercial software	Desktop forensic software for photo and video authentication, tampering analysis and source-device identification.	Specialist tool for trained examiners; best commissioned through a forensic expert.
Magnet Forensics magnetforensics.com Commercial software	Digital forensics software for acquiring and analysing data from phones, computers and cloud accounts.	Used by investigators and forensic examiners; relevant when a chat log or device record is in dispute.
TinEye tineye.com Web service	Reverse image search that finds earlier copies of an image online.	Helpful to show an image existed earlier elsewhere; cannot find images that were never indexed.

E-signature with audit trails

Tool	What it does	Keep in mind
DocuSign docusign.com Commercial	E-signature platform that produces a certificate of completion and audit trail for each envelope.	An audit trail shows the signing process; it does not by itself prove who the signer was or that the content is accurate.
Adobe Acrobat Sign adobe.com Commercial	E-signature service inside Adobe Acrobat with signing audit reports.	Same limits as other e-signature tools: it records the process, not the truth of the content.

Identity verification and KYC

Tool	What it does	Keep in mind
Entrust (formerly Onfido) onfido.com Commercial	Identity verification with document and biometric checks; Onfido is now part of Entrust.	Fraudsters now use AI-generated IDs and face injection, so identity checks should be layered, not relied on alone.

Tool	What it does	Keep in mind
Persona withpersona.com Commercial	Configurable identity verification and fraud-prevention platform.	Same layered-checks caveat as other KYC providers.
Jumio jumio.com Commercial	Identity verification, liveness detection and AML screening.	Same layered-checks caveat.
Veriff veriff.com Commercial	Identity verification with document, biometric and liveness checks.	Same layered-checks caveat.
InfoTrack infotrack.com.au Commercial	Australian provider of compliance (AML/CTF and identity), searches and workflow tools used by legal and property professionals.	Check which identity and AML products suit your obligations in your own jurisdiction.
First AML firstaml.com Commercial	AML and KYC onboarding software with regional versions (Australia, New Zealand, UK and global) used by law and conveyancing firms.	Same layered-checks caveat.

Citation checking and primary law

Tool	What it does	Keep in mind
Westlaw (KeyCite citator) legal.thomsonreuters.com Subscription	Thomson Reuters legal research platform; its KeyCite citator shows whether a case is still good law.	A citator checks existing authorities; it is the verification step for any authority produced by an AI tool.
Lexis+ (Shepard's citations) lexisnexis.com Subscription	LexisNexis research platform with Shepard's citation service and AI-assisted research.	Same use: confirm every authority in a primary source.
CoCounsel (Thomson Reuters) thomsonreuters.com Subscription	Thomson Reuters' AI assistant for legal research, drafting and document review.	Generative output must still be checked against the cited authority.
vLex (now part of Clio) vlex.com Subscription	Global legal research service with an AI assistant, now part of Clio.	Coverage varies by country; check it includes your jurisdiction.
Clearbrief clearbrief.com Commercial	Cite-checking tool for Microsoft Word that checks citations to law and facts and produces audit reports.	Checks what is in your draft; does not replace reading the authorities.
CourtListener citation lookup courtlistener.com Free (non-profit Free Law Project)	Free tool that finds citations in text and checks them against a database of US court opinions.	US case law only; absence from the database does not prove a case is fake.
BAILII bailii.org Free	Free British and Irish Legal Information Institute database of judgments.	Free access; coverage varies by court and period.
CanLII canlii.org Free	Free Canadian legal information database run by the Canadian Legal Information Institute.	Free access to Canadian decisions and legislation.

Tool	What it does	Keep in mind
AustLII austlii.edu.au Free	Free Australasian Legal Information Institute database of cases and legislation.	Free access; use for confirming that an Australian authority exists.

Tools at a glance

Every tool in this guide, by category

Tool	Category	What it is for	Access
Reality Defender	Deepfake and synthetic media detection	Detection platform for synthetic audio, video and images, with products aimed at calls and meetings as well as uploaded files.	Commercial (contact sales)
Hive (now branded Deploy)	Deepfake and synthetic media detection	Classifiers that flag AI-generated and deepfake images, video and audio, mainly used by trust-and-safety and enterprise teams.	Commercial
Sensity AI	Deepfake and synthetic media detection	Forensic-style deepfake detection for video, images and audio through a web app and API, with analyst reports.	Commercial (trial offered)
Intel FakeCatcher	Deepfake and synthetic media detection	Intel research project for real-time detection of synthetic face video, described on Intel's research pages.	Research / enterprise
TrueMedia.org	Deepfake and synthetic media detection	Non-partisan, open-source initiative whose platform helps people assess whether images, audio and video are manipulated.	Non-profit initiative (request access)
Pindrop	Deepfake and synthetic media detection	Voice security and deepfake detection for calls, virtual meetings and contact centres.	Commercial (enterprise)
Resemble AI Detect	Deepfake and synthetic media detection	Detection of synthetic audio, video and images, offered by an AI voice company, with an API and real-time options.	Commercial
Google SynthID	Deepfake and synthetic media detection	Google's watermarking approach for content made with its own generative tools, with a related detector.	Google DeepMind (watermarking research)
C2PA (Coalition for Content Provenance and Authenticity)	Content provenance	Standards body behind the C2PA specification for Content Credentials: signed metadata recording how a piece of media was made and edited.	Open standard
Content Credentials Verify	Content provenance	Free web tool that reads Content Credentials from a file and shows the recorded provenance, if any.	Free
Content Authenticity Initiative	Content provenance	Adobe-founded community promoting Content Credentials, with explainers and open-source tools.	Free information / community
Adobe Content Authenticity	Content provenance	Adobe's service for attaching and inspecting Content Credentials on creative work.	Adobe service
Truepic	Content provenance	Photo and video authentication at capture, mainly for inspections, insurance and similar workflows.	Commercial

Tool	Category	What it is for	Access
TimeBinder	Timestamping and hashing	Browser-based service that records a file's SHA-256 fingerprint with a time and anchors it to the Bitcoin blockchain.	First Pooled TimeBind free, no card; then from US$5
OpenTimestamps	Timestamping and hashing	Free, open-source protocol and tools that anchor file hashes to the Bitcoin blockchain; proofs can be checked without trusting a company.	Free, open source
RFC 3161 (Time-Stamp Protocol)	Timestamping and hashing	The IETF standard for trusted timestamp authorities (TSAs) that sign a hash together with a time.	Open standard
FreeTSA	Timestamping and hashing	Free RFC 3161 time stamp authority operated as a public service, with web form and command-line options.	Free
DigiStamp	Timestamping and hashing	Commercial RFC 3161 timestamp authority with an archive, pricing page and add-ins for common document software.	Commercial
OriginStamp	Timestamping and hashing	Provider of blockchain timestamping (web service and API) and an e-archive product called OriginVault.	Commercial
ExifTool	Forensics and metadata	Phil Harvey's command-line tool for reading and writing metadata in images, video, audio and many documents.	Free, open source
FotoForensics	Forensics and metadata	Web tool that runs error-level analysis and shows metadata to help look at possible edits to an image.	Free web tool
Forensically	Forensics and metadata	Browser-based photo forensics toolkit with clone detection, error-level analysis and metadata viewing.	Free web tool
InVID-WeVerify verification plugin	Forensics and metadata	Browser plugin built for journalists to verify images and video: keyframes, reverse search and metadata.	Free browser extension
Amped Authenticate	Forensics and metadata	Desktop forensic software for photo and video authentication, tampering analysis and source-device identification.	Commercial software
Magnet Forensics	Forensics and metadata	Digital forensics software for acquiring and analysing data from phones, computers and cloud accounts.	Commercial software
TinEye	Forensics and metadata	Reverse image search that finds earlier copies of an image online.	Web service
DocuSign	E-signature with audit trails	E-signature platform that produces a certificate of completion and audit trail for each envelope.	Commercial
Adobe Acrobat Sign	E-signature with audit trails	E-signature service inside Adobe Acrobat with signing audit reports.	Commercial
Entrust (formerly Onfido)	Identity verification and KYC	Identity verification with document and biometric checks; Onfido is now part of Entrust.	Commercial
Persona	Identity verification and KYC	Configurable identity verification and fraud-prevention platform.	Commercial
Jumio	Identity verification and KYC	Identity verification, liveness detection and AML screening.	Commercial

Tool	Category	What it is for	Access
Veriff	Identity verification and KYC	Identity verification with document, biometric and liveness checks.	Commercial
InfoTrack	Identity verification and KYC	Australian provider of compliance (AML/CTF and identity), searches and workflow tools used by legal and property professionals.	Commercial
First AML	Identity verification and KYC	AML and KYC onboarding software with regional versions (Australia, New Zealand, UK and global) used by law and conveyancing firms.	Commercial
Westlaw (KeyCite citator)	Citation checking and primary law	Thomson Reuters legal research platform; its KeyCite citator shows whether a case is still good law.	Subscription
Lexis+ (Shepard's citations)	Citation checking and primary law	LexisNexis research platform with Shepard's citation service and AI-assisted research.	Subscription
CoCounsel (Thomson Reuters)	Citation checking and primary law	Thomson Reuters' AI assistant for legal research, drafting and document review.	Subscription
vLex (now part of Clio)	Citation checking and primary law	Global legal research service with an AI assistant, now part of Clio.	Subscription
Clearbrief	Citation checking and primary law	Cite-checking tool for Microsoft Word that checks citations to law and facts and produces audit reports.	Commercial
CourtListener citation lookup	Citation checking and primary law	Free tool that finds citations in text and checks them against a database of US court opinions.	Free (non-profit Free Law Project)
BAILII	Citation checking and primary law	Free British and Irish Legal Information Institute database of judgments.	Free
CanLII	Citation checking and primary law	Free Canadian legal information database run by the Canadian Legal Information Institute.	Free
AustLII	Citation checking and primary law	Free Australasian Legal Information Institute database of cases and legislation.	Free

One-page firm checklist

Print this page and walk through it with your partners, your trust-account signatories and your IT provider.

Payments and instructions

- [ ] We verify every payment instruction and every change in bank details by calling a number already on file.
- [ ] Two authorised people approve new payees and transfers above our threshold.
- [ ] Our payment details are in the engagement letter, with a statement that they do not change by email.
- [ ] Multi-factor authentication is on for all email and practice-management accounts.

Documents and evidence

- [ ] We ask for native files with metadata, not screenshots or re-saved PDFs.
- [ ] We hash (SHA-256) and independently timestamp key files at intake.
- [ ] We keep an untouched original and work on copies.
- [ ] We keep a chain-of-custody log, and re-hash when files move.
- [ ] Disputed images, video, audio and chats go to a forensic examiner when stakes are high.

AI use and citations

- [ ] We have a written AI policy: approved tools, confidentiality limits, named responsible lawyer.
- [ ] Every authority is checked in a primary source and a citator before filing.
- [ ] We log who verified each citation.
- [ ] We ask experts whether AI was used and check every source and credential.

People and identity

- [ ] We use layered identity checks, with liveness and source-of-funds questions for higher-risk clients.
- [ ] We re-verify identity when instructions or contact details change.
- [ ] We do not act on video or voice alone.

Governance

- [ ] Staff are trained on deepfake and payment-redirection scams at least annually.
- [ ] We have a written incident plan: stop, preserve, escalate, report.
- [ ] Our professional indemnity and cyber cover has been checked against these risks.
- [ ] This checklist is reviewed twice a year.

Client checklist: how to protect yourself when dealing with us

Give this to clients with the engagement letter. It is written in plain English and can be adapted.

Before you pay or share anything

- [ ] Our bank details will not change by email, text or phone message. If you receive such a message, call us on the number in your engagement letter before you pay.
- [ ] Never act on a payment request that arrives by email alone, even if it looks like it is from us.
- [ ] Be wary of urgent or secret requests, even on a video call. Hang up and call us back.

When you give us documents and evidence

- [ ] Send original files, not screenshots or photographs of a screen. Do not edit, rename, re-save or "tidy up" them first.
- [ ] Tell us where and when each file was created, on which device, and who has had it.
- [ ] Keep the original device, account or backup safe. Do not delete anything related to the matter.
- [ ] If something you received looks wrong or you suspect it was edited, tell us before you rely on it.

Keeping your accounts safe

- [ ] Use a different strong password for your email and turn on two-step verification.
- [ ] Tell us immediately if your email or phone has been hacked, or you have been contacted by someone claiming to be us.

Sample engagement-letter clause (template to adapt)

TEMPLATE ONLY

The text below is a starting point for a firm's lawyer to adapt. It is not legal advice, it is not drafted for any particular jurisdiction, and it must be reviewed against your own regulator's rules, client-care requirements and insurance terms before use.

PAYMENT AND INSTRUCTION VERIFICATION

We will never ask you to pay money to a new or changed bank account by email, text message or voicemail alone. Our payment details are as set out in this letter. If you receive any message that appears to come from us and gives different payment details, do not pay. Call us on [firm main number] before doing anything.

We may decline to act on any instruction to pay or transfer money that we have not been able to confirm by calling you on a number we already hold. Voice or video contact alone is not sufficient confirmation. You agree to cooperate with those checks, and understand that they may cause delay.

DOCUMENTS, EVIDENCE AND RECORDS

When you give us documents, photographs, messages or recordings, please provide the original files in their original (native) format with their metadata intact, and tell us how, when and on what device they were created. Please do not edit, rename, re-save or forward them first. We may create a digital fingerprint (hash value) of files you provide, and may record an independent timestamp of that fingerprint, so that we can later show that a file has not changed since we received it. [Where applicable: Fingerprinting is done in a way that does not transmit the content of your files to any third party.]

You agree to tell us promptly if you believe any document or evidence has been altered or created using artificial intelligence, or if you suspect you have been the target of fraud or impersonation.

USE OF TECHNOLOGY

We may use artificial intelligence tools to assist with some tasks. A lawyer remains responsible for our work, and we verify authorities and factual statements against primary sources before relying on them. [Describe confidentiality arrangements.]

Sample evidence-intake protocol (template to adapt)

TEMPLATE ONLY

Adapt this procedure to your practice area, systems and jurisdiction. It is not legal advice. Where evidence may be tendered in court, follow the applicable rules and any expert guidance.

How to receive and preserve digital evidence

1. **Log the receipt.** Record who sent it, how (channel), when, who received it, and any intermediate handlers. Use one intake log per matter.
2. **Ask for native originals.** Request the original file from the original device or account, in its native format, not a screenshot or re-saved copy. Record any reason it cannot be provided.
3. **Do not open or edit the original.** Store it read-only. All review, annotation and redaction is done on a working copy.
4. **Create a SHA-256 hash of the original as received.** Record the hash value in the intake log alongside the filename, size and time.
5. **Timestamp the hash independently.** Use an RFC 3161 authority or a blockchain-anchored service, and store the proof or certificate with the file.
6. **Capture and store metadata.** Export metadata with a tool such as ExifTool and store the output next to the original. Treat it as a claim, not as proof.
7. **Ask provenance questions.** Who created it, with what device and app, when, where has it been stored, who else has had access, and has it been edited or converted? Record the answers.
8. **Triage for red flags.** Check against the red flags in this guide. For images, video and audio consider running a detector and a reverse-image search, and record the results as signals only.
9. **Corroborate.** Find at least one independent source for key facts and dates: a call-back to the issuer, platform logs, third-party records.
10. **Escalate where needed.** If a high-stakes item looks doubtful, stop, tell the supervising partner and consider a forensic examiner. Do not tender anything you doubt.
11. **Maintain the chain of custody.** Log every access and transfer. Re-hash after any move and confirm the hash still matches.

If you suspect a forgery: a short response plan

1. **Stop.** Do not pay, file, serve or forward the item.
2. **Preserve.** Keep the original file, headers, logs and messages. Do not alter or "clean" anything. Hash and timestamp what you hold.
3. **Escalate.** Tell the supervising partner and the person responsible for risk or compliance.

4. **Verify independently.** Use call-backs, primary sources and, where warranted, a forensic expert.
5. **Recover and report.** For payment fraud, contact your bank at once and the relevant fraud or cybercrime authority (FBI IC3 in the US, Scamwatch in Australia, or your local equivalent).
6. **Consider your duties.** Think about duties to the client, to the court and to your insurer and regulator, and take advice. If a false item has already been filed, correct it promptly and candidly.
7. **Learn.** Record what happened and update the procedures and the checklist.

Frequently asked questions

How can lawyers prevent AI forgeries?

No single measure prevents them. Lawyers combine out-of-band verification of people and payment instructions, native-file evidence handling with chain of custody, hashing and independent timestamping at intake, checking every authority in a primary source, layered identity checks, and a clear engagement-letter clause and client checklist.

Can AI-detection tools prove a document or video is fake?

No. Detection tools return probabilities, can miss new generators and can flag genuine material. Use them as one signal, record the result, and use a qualified forensic examiner for high-stakes matters.

How do you prove a document was not backdated?

Metadata and file dates can be edited, so they are weak evidence on their own. An independent timestamp of the file's hash shows the exact file existed no later than a recorded time, and corroboration from third-party records adds weight. A timestamp does not show who created the document or that its contents are true.

What is the best defence against deepfake CEO or partner payment scams?

An out-of-band call-back to a number you already hold, plus a two-person approval rule for new payees and large transfers. Realistic voice or video is no longer proof of identity.

How do I avoid citing fake AI cases in court?

Check every authority in a primary source before it leaves the firm: read the case, confirm the quote and run a citator. Keep a log of who verified each citation. Cases from the United States, England and Wales, Canada and Australia show the consequences of not doing so.

Is a timestamp or blockchain record admissible in court?

That depends on the jurisdiction, the court and the case. A timestamp is one piece of evidence about when an exact file existed; it should be considered with other evidence and with legal advice. Some rules, such as Rule 902(14) of the US Federal Rules of Evidence, allow authentication of copied data by hash value through a qualified person's certification.

What are Content Credentials and C2PA?

Content Credentials are signed provenance records, built on the open C2PA standard, that can record how a piece of media was created and edited. They help when present and valid, but are often missing or stripped, so a lack of credentials does not show a file is fake.

What should be in an engagement letter to protect against payment fraud?

A statement that your bank details will not change by email, a requirement that clients call you on a known number before paying any changed details, a note that voice or video alone will not confirm payment instructions, and a request that clients preserve and provide evidence in native form.

Do these steps work in every country?

The practices are general and apply broadly across the US, UK and Europe, Australia and Canada. The legal rules on evidence, privilege, professional conduct and identity checks differ, so check your own regulator and court rules.

Glossary

AI forgery
A false or altered document, image, audio, video, message or authority created or modified with generative AI to look genuine.

Audit trail
A record of the steps in a process, such as who was sent, opened and signed an e-signature envelope, and when.

Backdating
Creating or altering a document so it appears to date from an earlier time than it really does.

Blockchain anchoring
Recording a hash in a public blockchain so that its existence at that time can be verified later.

Business email compromise (BEC)
Fraud that uses a compromised or spoofed email account to trick someone into paying money or sharing data.

C2PA
The Coalition for Content Provenance and Authenticity, which publishes an open standard for provenance data called Content Credentials.

Call-back verification
Confirming an instruction by calling a number you already hold, not a number supplied in the message.

Chain of custody
A written record of who handled evidence, when, where it was stored and what was done to it.

Content Credentials
Signed metadata attached to media recording how it was created and edited, built on the C2PA standard.

Deepfake
Synthetic audio, video or images that realistically imitate a real person or event.

EXIF metadata
Camera and file information embedded in photos, such as device, settings and date. It can be edited.

Hallucination
When an AI model generates plausible but false information, such as a case that does not exist.

Hash (fingerprint)
A short string computed from a file's exact contents. Any change to the file changes the hash.

KYC
"Know your customer": identity checks a business performs on clients.

Liveness detection
Checks that a person is physically present during an identity verification, rather than a photo, recording or deepfake.

Native format
A file in the format created by the original software, with its metadata, rather than a screenshot or printout.

Provenance
The documented history of where a file came from and how it changed.

RFC 3161
The internet standard for trusted timestamp authorities, which sign a hash together with a time.

SHA-256
A widely used hash algorithm that produces a 256-bit fingerprint.

Time stamp authority (TSA)
A third party that issues signed timestamps for hashes, usually under RFC 3161.

Voice cloning
Using AI to generate speech that imitates a particular person's voice.

Sources and further reading

Incidents and facts in this guide come from the following sources, all opened and checked on 11 November 2026. If a link changes or a case is appealed, rely on the primary source.

- CNN: Arup revealed as victim of US$25 million deepfake scam (16 May 2024)
- The Guardian: UK engineering firm Arup falls victim to £20m deepfake scam (17 May 2024)
- Wall Street Journal: Fraudsters used AI to mimic CEO's voice (2019)

- Mata v. Avianca, Inc., opinion and order on sanctions (S.D.N.Y., 22 June 2023)
- Ayinde v London Borough of Haringey; Al-Haroun v Qatar National Bank [2025] EWHC 1383 (Admin)
- Zhang v Chen, 2024 BCSC 285
- Victorian Legal Services Board: statement on the Dayal matter (2 September 2025)
- Mendones v. Cushman & Wakefield, order re terminating sanctions (Cal. Super. Ct., 9 September 2025)
- Kohls v. Ellison, order on expert testimony (D. Minn., January 2025)
- FBI Internet Crime Complaint Center: 2025 Internet Crime Report
- 404 Media: Inside the underground site where AI churns out fake IDs
- BleepingComputer: OnlyFake operator pleads guilty
- US Federal Rules of Evidence, Rule 902 (including 902(14))
- American Bar Association: Formal Opinion 512 on generative AI (July 2024)
- The Law Society (England and Wales): generative AI, the essentials

Disclaimer. This guide is general information for lawyers and is not legal advice. Tool descriptions summarise vendor claims and are not endorsements. Laws, rules, products and prices change; check current sources before relying on anything here.