

PAPER THREE OF THREE • LEGAL SERVICES

Verified Authority

Artificial intelligence in law firms, and the three duties it has already been found to threaten

Paul Forrest

Fractional Head of AI, Ecaveo
September 2026

About this paper

This is the third of three papers on deploying artificial intelligence inside professional firms that carry personal accountability for their work. The first concerns private equity, the second accounting and audit firms. Each paper stands alone. Together they argue one thing, which is that the durable advantage in professional work comes from decisions that can be reconstructed afterwards, and that systems which produce answers without producing a record destroy more value than they create.

Law comes last in the series because it is the only one of the three where the failure has already been litigated. A Divisional Court has considered fabricated authorities twice in a single judgment, a tribunal has addressed privilege directly, and a regulator issued a warning notice in August 2026 against a backdrop of live investigations.^{5,9,10} The profession is past the stage of anticipating what might go wrong.

Editorial balance. Roughly seven parts in ten of this document is evidence and sector analysis, two parts in ten is original framework, and one part in ten concerns how I would run the work. Where the evidence is thin, the paper says so.

HOW TO READ THE EVIDENCE TAGS

Every substantive claim carries a tag stating what kind of evidence supports it.

REGULATOR A regulator or professional body speaking in its own voice.

CASE LAW A judgment or tribunal decision.

STANDARD A published standard, framework or professional requirement.

ADVISORY An advisory opinion or consultation from a body with no power to bind.

PEER-REVIEWED A primary study, working paper or systematic review.

VENDOR RESEARCH Research or a survey from a firm with a commercial interest in the answer. Directional only.

PRACTICE Practitioner opinion, including my own.

EVIDENCE GAP A question the available evidence does not answer.

ECAVEO My own framework, proposed rather than proven.

Legal and regulatory notice. This paper describes regulatory guidance, professional obligations and decided cases throughout, and Chapters 2 and 5 in particular. Nothing in this document constitutes legal advice. I am not a lawyer. The provision of legal advice sits outside the terms of any engagement with the author or with Ecaveo, and the material is presented to support discussion and further review by qualified advisers.

Statements about what a source says are attributed to that source. Where I draw a practical implication the source does not itself state, the text says so. Guidance and case law in this area are moving quickly.

Contents

A note from the author 4

Executive summary 5

ONE The duty that does not move 7

TWO What the regulator and the courts have said 10

THREE Where the value sits across a firm 14

FOUR Six ways this fails on a matter 17

FIVE Privilege, and the control model 20

SIX Proving it, and the evidence gate 24

SEVEN Verification, economics and the fee 26

EIGHT Ninety days, then nine months 28

NINE The session to run with the partners 29

TEN Where the evidence is thin 30

References 32

THE ARGUMENT IN ONE PARAGRAPH

Artificial intelligence is useful in a law firm wherever an authoritative source exists to check the output against, and dangerous wherever one does not. The regulator has said that solicitors remain accountable for their work and outputs regardless of how that work was prepared.⁵ A tribunal has said that uploading confidential documents to an open tool places them in the public domain.¹⁰ Neither statement asks a firm to do anything new. Both require a firm to show that it did the thing it has always been required to do, which means verification has to become a recorded step. Assuming it happened is what the cases turn on.

ILLUSTRATIVE MATERIAL

Every firm, client, matter and person in the worked examples is invented. Each is marked as an illustrative scenario where it appears. Decided cases are cited by neutral citation and described from the judgment. Invented material is never presented as evidence.

CITATION

Forrest, P. (2026). *Verified Authority: Artificial Intelligence, Privilege and the Duty to the Court*. Ecaveo Responsible AI Series, Paper Three. Version 1.0, September 2026.

ALSO IN THIS SERIES

Paper one. *The Traceable Thesis*. Artificial intelligence across the private equity investment lifecycle.

Paper two. *Sufficient and Appropriate*. Artificial intelligence, audit evidence and professional scepticism.

A NOTE FROM THE AUTHOR

Why this paper is the shortest on opportunity

Every other sector in this series is arguing about where the value might be. Law is the only one where the cost of getting it wrong has already been published, with names attached.

I came to this through business process work in the 1990s and moved into machine learning and language models from 2014. I build small, specific models in preference to large general ones, because a model trained on everything is calibrated for nothing, and legal work punishes confident approximation more than most work does.

I am not a lawyer, and this paper is written as a board-level adviser would write it. That shapes what it does. It reports what regulators and courts have said, grades how firmly they said it, and keeps my own inferences visibly apart from theirs. Where a point bears on a firm's own exposure, the text says that a lawyer should confirm it.

What prompted the paper was the tone of the market. Legal technology is being sold on time saved, and the published record now contains a Divisional Court judgment concerning fabricated authorities in two separate matters, a tribunal decision addressing privilege in terms, and a regulatory warning notice issued while investigations were live.^{5,9,10} Those are not edge cases. They are the first sample of what happens when a professional relies on an output nobody checked.

The argument here is narrow and I think it holds. The value in a law firm sits wherever an authoritative source exists to check against, and the risk sits wherever one does not. A firm that organises its deployment around that single distinction picks up most of the benefit. The exposure sits almost entirely on the other side of the line.

TWO PAGES THAT CIRCULATE ON THEIR OWN

Executive summary

The regulator has spoken, the courts have ruled, and the measured reliability of the leading tools is lower than the market implies. All three facts are recent and all three are public.

Three duties are engaged by this technology, and each has now been considered by an authority. The duty not to mislead the court has been considered by a Divisional Court. The duty of confidentiality, and with it privilege, has been considered by a tribunal. The duty to remain accountable for one's own work has been stated by the regulator. Nothing in this area now rests on speculation.

1

The regulator has stated that accountability does not transfer. The warning notice of 17 August 2026 states that artificial intelligence has no separate legal personality and that solicitors and regulated individuals remain accountable for their work and outputs regardless of how that work has been prepared.⁵ **REGULATOR** It requires that submissions of named case law authorities are genuine, relevant and carry a verifiable citation.

2

The courts have already dealt with fabricated authority, and the sanctions were regulatory. In June 2025 a Divisional Court considered two matters together, one involving five fabricated housing authorities and one involving 45 problematic citations of which 18 were entirely fictitious. Wasted costs of £2,000 were ordered in the first, and five referrals followed covering the practitioners and their firms.⁹ **CASE LAW**

3

A tribunal has addressed privilege in open tools directly. An Upper Tribunal decision reported in the 2026 series held that uploading confidential documents into an open tool places the information on the internet in the public domain, breaching confidentiality and waiving privilege, and drew a distinction between open and closed systems.¹⁰ **CASE LAW** The Upper Tribunal is a superior court of record by statute, and a reported decision of its Immigration and Asylum Chamber binds the First-tier Tribunal. It does not bind the High Court or above.

4

The measured reliability of the leading legal research tools is well short of the marketing. An academic study of 202 constructed queries reported hallucination rates of 17 per cent and 33 per cent for two leading tools, against a range of 58 to 82 per cent reported for general-purpose models on separate material.¹³ **PEER-REVIEWED** Both vendors disputed the study, one on product selection and one on the definition used.

5

The problem is large, international and mostly not committed by lawyers. A public tracker recorded 2,036 cases involving fabricated citations as at 10 September 2026, of which 69 were in the United Kingdom, and rather more involved self-represented litigants than lawyers.¹² **SURVEY** The compiler states that the collection seeks to be exhaustive and remains a work in progress, so it should be read as an order of magnitude.

6

The fee question is live in the United States and unaddressed here. A formal ethics opinion states that a lawyer may not bill for time saved and that a fee for which little or no work was performed is unreasonable.¹¹ **ADVISORY** I could find no equivalent statement from the regulator or the representative body in England and Wales. **EVIDENCE GAP**

What follows for a firm

The deployment principle is one line. Use these tools where an authoritative source exists to check the output against, and do not use them where one does not. Research against an approved primary corpus qualifies. Clause extraction from a permissioned document set qualifies. A confident answer about the law from a general model does not, and the measured reliability figures explain why.

Underneath that sits a records discipline. Verification should be a step in the workflow with a name attached. An instruction to be careful is not a control. A matter file should show which authorities were checked, against what, and by whom. That record costs minutes and is the only thing that answers a regulator or a court.

The third thing is a decision about tools. An open consumer service and an enterprise service under contract are different objects, and a tribunal has now said so in terms.¹⁰ A firm that has not drawn that line clearly, in writing, and enforced it technically, is relying on the judgement of every fee earner every day.

Chapter 8 sets out a 90-day proof on three bounded workflows. Chapter 10 states what would change the position taken here.

CHAPTER ONE

The duty that does not move

The obligations engaged here are old. What is new is that a fluent draft now arrives before anybody has decided whether it is right.

A solicitor owes duties of competence, of confidentiality, of supervision and of candour to the court. None of them contains an exception for work produced with assistance. The regulator has said as much in terms, stating that artificial intelligence has no separate legal personality and that regulated individuals remain accountable for their work and outputs regardless of how that work has been prepared.⁵ **REGULATOR**

The reason this needs saying at all is a change in the order of events. A lawyer researching a point used to arrive at a proposition after reading the authorities, so verification and formation happened together. A lawyer using a general model arrives at a proposition first, in polished prose, with citations attached, and verification becomes a separate act that somebody has to remember to perform. The duty is unchanged and the workflow that used to discharge it automatically has gone.

That is the whole of the problem, and it explains why the failures that have reached the courts look careless. The Divisional Court in June 2025 found no evidence that one practitioner had set out to deceive, treating the matter as a failure to verify research supplied by the client.⁹ **CASE LAW**

THE CORE CLAIM

Verification used to be a by-product of doing the work. It is now a separate step, and a step that is not designed in will not happen.

1.1 The cycle this series uses

Every paper in this series describes the same five-stage cycle. In legal work the second stage is where professional obligation actually bites, and it is the stage the current generation of tools is least willing to perform.

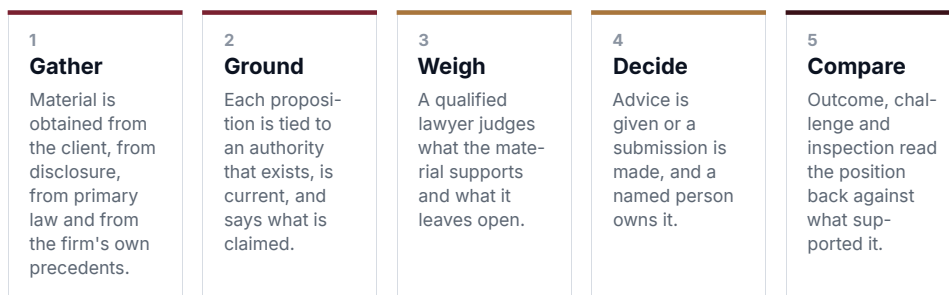
THE CYCLE EVERY PAPER IN THIS SERIES EXAMINES

FIGURE 1 Stage two is the one the profession has always performed silently. A tool that produces stage one output at stage four quality is a tool that invites stage two to be skipped.

Ecaveo framework, used across the series. **ECAVEO**

1.2 How reliable the tools actually are

There is measured evidence here, which is unusual in this field, and it deserves to be quoted accurately. A study of 202 hand-constructed legal queries across general research, jurisdiction and time-specific questions, false-premise questions and factual recall, reported that one leading tool produced accurate grounded answers in 65 per cent of cases with a hallucination rate of 17 per cent, and that a second produced accurate answers in 41 per cent of cases with a hallucination rate of 33 per cent.¹³ **PEER-REVIEWED** The same authors cite earlier work measuring hallucination by general-purpose models on legal questions at between 58 and 82 per cent, on a separate set of queries.¹³ **PEER-REVIEWED**

Both vendors disputed the original work. One argued that the study had tested a product limited to secondary-source content, not the research product intended for primary law, access to which had been refused. The other disputed the definition of hallucination as too broad and reported lower rates in internal testing.¹³ **VENDOR RESEARCH** The first objection was then answered. Access was granted, the study was extended to cover the research product in June 2024, and the peer-reviewed version followed in 2025. The 33 per cent figure quoted above comes from that extension.

Purpose-built legal tools hallucinate far less than general models and far more than a lawyer relying on them would assume.

Two conclusions survive the dispute. Purpose-built tools perform substantially better than general-purpose models on legal questions, which is an argument for procurement. None of the tools measured performed well enough to remove the verification step, which is an argument about workflow. My own view is that the second conclusion is the operationally important one, and it does not depend on which vendor is right about the first.

1.3 Adoption, and the quality of the numbers

Adoption figures in this market come almost entirely from vendors, and a partner reading them should know that. A survey conducted in January 2025 reported that 46 per cent of UK legal professionals used these tools for work, against 11 per cent in July 2023, although the sampling frame was not published.¹⁴ **VENDOR RESEARCH** A later survey of 543 UK legal professionals reported 94 per cent using them, with 83 per cent expressing concern about fabricated or inaccurate output.¹⁵ **VENDOR RESEARCH**

One figure cuts against the commercial interest of the firm that published it, and is more informative for that reason. A 2025 study of US legal professionals reported that while 79 per cent used artificial intelligence, the proportion using legal-specific tools had fallen from 58 per cent to 40 per cent.¹⁶ **VENDOR RESEARCH** Read at face value, that says the profession is increasingly using general tools for legal work, which is precisely the pattern the reliability evidence argues against.

The regulator noticed the direction of travel early. Its risk outlook of November 2023 reported that around three quarters of large firms were using some form of artificial intelligence by the end of 2022, that more than 60 per cent were exploring generative tools, and that lawyers had already submitted non-existent cases.⁶ **REGULATOR**



Transactional review. Extraction against a permissioned document set is the safest large-scale use in a firm, because the source is present and the check is mechanical.

CHAPTER TWO

What the regulator and the courts have said

Four authorities matter and they say compatible things. Two are decisions of a court, one is regulatory guidance and one is guidance to judges.

A reminder on register. This chapter summarises what published sources say and grades how firmly they say it. I am not a lawyer and this is not a statement of the law. Where I draw an implication a source does not itself state, the sentence says so, and any point bearing on a firm's own position should be confirmed by counsel.

2.1 The warning notice

On 17 August 2026 the Solicitors Regulation Authority issued a warning notice on the misuse of artificial intelligence.⁵ **REGULATOR** Four elements of it bear on firm policy.

It describes the technical failure plainly, stating that these tools can produce hallucinations, generating fictitious cases, references or seemingly factual assertions that may appear convincing despite having no basis in fact. It addresses the duty to the court, requiring that representations be properly arguable and that submissions of named case law authorities are genuine, relevant and carry a verifiable citation. On confidentiality it states that information entered into these systems may be stored, retained or used to improve the tool, creating risks to confidentiality, legal professional privilege and compliance with data protection law. On accountability it states that these systems have no separate legal personality and that regulated individuals remain accountable for their work and outputs regardless of how that work has been prepared.⁵ **REGULATOR**

Worth highlighting is what the notice does not do. It sets no threshold for adequate verification, prescribes no tooling, and offers no safe harbour for a firm that follows a particular process. My own view, which the notice does not state, is that the absence of a prescribed standard makes a firm's own documented standard the thing that will be examined, and that a firm without one will be judged against what a competent firm should have done.

REGULATOR

Contemporaneous trade reporting indicates the notice was issued while the regulator had investigations into misuse open. A firm should read it as active.

2.2 The judgment on fabricated authority

In June 2025 a Divisional Court considered two matters together under the jurisdiction the court holds over those who appear before it.⁹ **CASE LAW** In the first, counsel had cited five fabricated housing authorities in grounds for judicial review and had also misstated a provision of the Housing Act 1996. In the second, a solicitor had put forward 45 problematic citations of which 18 were entirely fictitious, having failed to verify research supplied by the client.

The outcomes were regulatory, and this is frequently misreported in both directions. Wasted costs of £2,000 were ordered against counsel and against the instructing law centre in the first matter. The court held that the threshold for contempt was arguably met and declined to initiate proceedings as a matter of discretion, giving as its first reason that a number of factual issues could not easily be determined in summary contempt proceedings, and noting that counsel had already been publicly criticised and referred to her regulator. It raised questions as to potential failings by those responsible for her training, supervision, pupillage sign-off, work allocation and marketing, which is a question raised and no more than that. In the second matter it found no evidence of deliberate deception.⁹ **CASE LAW**

The referrals went wider than the two practitioners. Counsel was referred to the Bar Standards Board. The instructing solicitor and the law centre in the first matter, and the solicitor and his firm in the second, were referred to the Solicitors Regulation Authority, which makes five referrals covering firms as well as individuals.⁹ **CASE LAW**

The statement of principle is the part that will be cited for years. Adopting the words of Dias J, who had referred the second matter to the court, the Divisional Court recorded that the administration of justice depends upon the court being able to rely without question on the integrity of those who appear before it and on their professionalism in only making submissions which can properly be supported.⁹ **CASE LAW** A firm building policy should treat that sentence as the governing one, and the second limb as the operative half, because a submission that can be properly supported is a submission somebody has checked.

2.3 The decision on privilege

A decision of the Upper Tribunal reported in the 2026 series, and promulgated in November 2025, addressed the confidentiality question in terms that no earlier English authority had.¹⁰ **CASE LAW** It held that uploading confidential documents into an open-source tool such as a public chatbot is to place that information on the internet in the public domain, and thus to breach client confidentiality and waive legal privilege. The tribunal distinguished open public models from closed or enterprise tools.

One qualification should travel with that holding whenever it is cited. The Upper Tribunal is a superior court of record and a reported decision of its Immigration and Asylum Chamber binds the First-tier Tribunal, and it does not bind the High Court or the Court of Appeal, so a higher court could reason differently. The analysis has since been echoed in extra-judicial commentary and in a broad consensus among firms advising on the point, which strengthens confidence without changing on whom the decision binds. On my reading, a firm would be unwise to rely on any argument that privilege survives disclosure to an open consumer service, and that is a point on which a firm should take its own advice rather than mine.

The same decision extended supervisory liability in a direction that matters for training. It observed that a supervisor who fails to prevent a junior fee earner's false citations is likely to be more culpable than one who fails to catch their own errors.¹⁰ **CASE LAW**

THE FOUR AUTHORITIES AT A GLANCE

SOURCE	WHAT IT SAYS	WEIGHT
SRA warning notice, 17 August 2026	Accountability does not transfer; authorities must be verifiable, and inputs may be retained ⁵	Regulatory guidance, enforceable through the conduct rules
Divisional Court, June 2025	Fabricated authority attracts wasted costs and regulatory referral, and the court must be able to rely on those before it ⁹	Judgment of a court of record
Upper Tribunal, 2026 series	Uploading to an open tool breaches confidentiality and waives privilege; open and closed systems differ ¹⁰	Superior court of record. Binds the First-tier Tribunal, and does not bind the High Court or above
Judicial guidance, October 2025	Accuracy must be checked, nothing confidential entered into public chatbots, and office holders are personally responsible ⁷	Guidance to judges, indicative of judicial expectation

TABLE 1 Read down the right-hand column. Two of these are judicial decisions and two are guidance, which affects how firmly each can be relied on and on whom each binds.

2.4 The judiciary, and what it watches for

The guidance to judicial office holders issued in October 2025, the third iteration after versions in December 2023 and April 2025, states that the accuracy of output must be checked before use, that confidential or private information should never be entered into public chatbots because anything typed in could become publicly known, and that judicial office holders are personally responsible for material produced in their name.⁷ **REGULATOR**

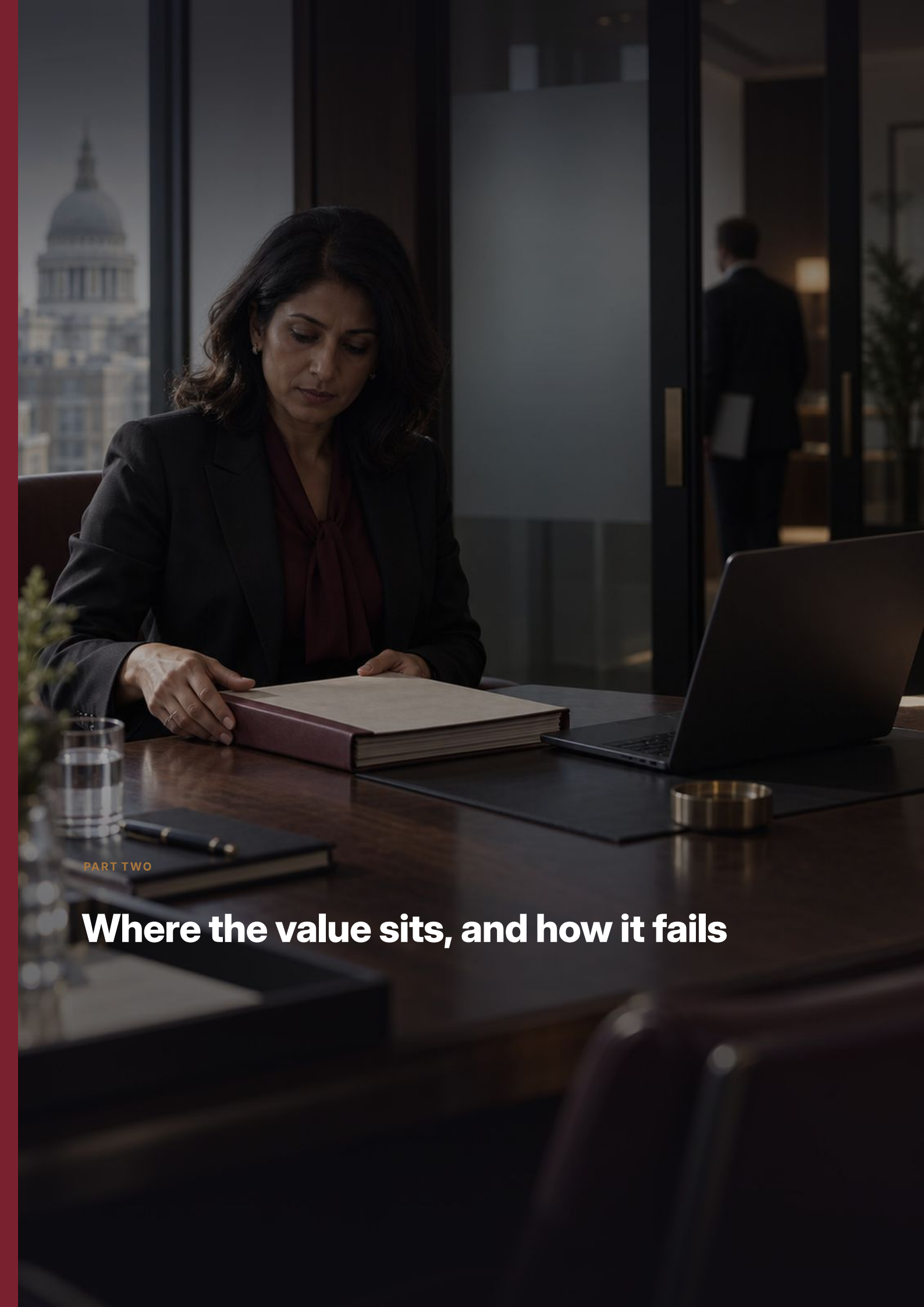
It also tells judges what to look for. Unfamiliar citations, citations in an American style, American spelling, all of it is identified as warranting scrutiny.⁷ **REGULATOR** A firm should take that as a practical warning. Submissions are now being read by people who have been told how to spot unverified work.

2.5 What is still open

The Civil Justice Council consulted in February 2026 on the use of these tools in preparing court documents, proposing a tiered approach.⁸ **ADVISORY** Its interim position was that statements of case and skeleton arguments need no declaration, because existing professional responsibility rules suffice where a representative's name is on the document. It proposed an express declaration for trial witness statements under the relevant practice direction, to the effect that the technology was not used to alter, embellish, strengthen, dilute or rephrase the witness's own evidence, and proposed that experts explain the tools used beyond purely administrative functions.

An update published in July 2026 reported an emerging consensus to maintain the position for professional legal drafting, continued disagreement on witness statement safeguards, movement towards proportionate transparency for expert evidence, and litigants in person as an unresolved question.⁸ **ADVISORY** The final report was expected later in 2026 and had not appeared as at the date of this paper.

The representative body's own guidance, dated September 2025, covers the practical ground. It warns that providers may see input and output data and that data often leaves the United Kingdom, stresses authentication, notes that vendors may retain intellectual property in outputs, and recommends telling clients when these tools are used on their matter.¹⁷ **REGULATOR**



PART TWO

Where the value sits, and how it fails

CHAPTER THREE

Where the value sits across a firm

Order the portfolio by one question. Does an authoritative source exist against which the output can be checked cheaply?

That single test sorts almost every proposed use in a law firm, and it sorts them better than any taxonomy based on practice area. Where an authority exists and the check is quick, the tool pays. Where the check requires redoing the work, it does not. Where nothing exists to check against, the firm is relying on the model, which the reliability evidence says it should not do.

THE OPPORTUNITY PORTFOLIO

WORKFLOW	WHAT THE TECHNOLOGY DOES WELL	THE VALUE TEST
Document review	Extracts clauses, compares positions, builds chronologies and finds anomalies	Coverage against a known answer set, with lawyer review of every conclusion
Legal research	Maps issues and retrieves authority from approved, current and citable sources	Breadth of first-pass coverage, with primary-source verification mandatory
Drafting and negotiation	Produces first drafts, clause options, issue lists and explanations of redlines	Cycle time, with lawyers owning advice, strategy and final language
Litigation preparation	Organises evidence and builds source-linked chronologies from a controlled bundle	Event recall, date accuracy and the correction effort a lawyer has to apply
Knowledge and operations	Retrieves precedents and policies, summarises calls and captures lessons	Reuse of firm knowledge, subject to ethical walls and matter permissions
Matter intake	Classifies enquiries, extracts facts and identifies missing information	Triage speed, with acceptance, conflicts and risk decisions kept separate

TABLE 2 The value test in place of the benefit. A benefit measured against nothing will not survive a supervising partner.

3.1 Document review is the strongest case in the firm

Review has the properties that make evaluation possible and verification cheap. A closed matter has a known answer set, the source document is present, and checking an extracted clause means opening the clause. A firm can take three completed transactions, rebuild the document sets, and measure recall, false positives and review time against what the team actually found.

The regulator's own risk outlook recorded early examples of review time falling from weeks to seconds, and a claims-processing implementation reporting a 77 per cent cost reduction.⁶ **REGULATOR** Those are reported figures, and a firm should run its own test before adopting them. My experience is that the realised gain concentrates in coverage, because teams use the capacity to look at more.

ILLUSTRATIVE SCENARIO

The chronology that was internally consistent

An invented litigation team builds a chronology from a 9,000-document disclosure set. The output is dated, source-linked and internally consistent, and it becomes the working spine of the case.

Two months later a supplemental disclosure produces a document from the middle of the relevant period which contradicts the sequence the chronology established. The original set had been loaded before a custodian's archive was collected. The tool had reported accurately on what it held.

The chronology was never wrong about its own contents, and it said nothing at all about what it did not have. The set ran to 9,000 documents, and nobody asked what the 9,001st would have said.

3.2 Research, and the corpus that makes it safe

Research is where the reliability evidence bites hardest, and where design makes the largest difference. A general model asked a question of law will answer it. A system restricted to an approved primary corpus, required to cite, and permitted to decline, answers a different and safer question, which is what the approved corpus says.

Three acceptance criteria follow, and I would apply all three before a research tool reaches a fee earner. Every proposition carries a citation that resolves to a document in the corpus. Jurisdiction and currency are stated, since an authority that has been overruled is worse than no authority. The system declines when the corpus does not support an answer, and the rate of appropriate refusal is measured. **ECAVEO**

PRACTICE

Treat appropriate refusal as a feature to be tested. A research tool that never declines is a research tool that will one day invent.

3.3 Drafting, where the benefit is real and the risk is diffuse

First drafts, clause options and explanations of a counterparty's redline are genuinely faster to produce than to write. The risk here is subtler than fabrication. A generated clause reads as though it has been considered, and a reviewer reading fluent text applies less scrutiny than a reviewer reading a colleague's rough draft.

The practical control is a house rule. Generated language is marked as generated until a lawyer has adopted it, and adoption is an act somebody performs. Saying nothing does not count. Firms that have adopted this tell me it changes reviewer behaviour more than any training did, though I have no measurement to offer and neither did they. **PRACTICE**

3.4 What to leave alone for now

Two categories should stay out of the first wave. Anything that produces a filed document without a named lawyer verifying every authority in it belongs there, for reasons the judgment in Chapter 2 makes obvious. Anything that touches conflicts or information barriers belongs there, because those are deterministic controls and a retrieval system is a probabilistic one.

Neither is permanent. Both describe work where the cost of an error is high, the ability to detect it is low, and the profession's own published record shows what happens when it goes wrong.



Evidence and chronology. A source-linked chronology is only as good as its account of what it does not contain.

CHAPTER FOUR

Six ways this fails on a matter

Two of these six have already produced published consequences for named practitioners. The other four have not yet, which is a statement about detection.

THE FAILURE TAXONOMY

FAILURE	WHAT ACTUALLY HAPPENS	DETECTED IN PRACTICE
Fabricated authority	A case, statute or quotation that does not exist enters advice or a filing	Yes, by opponents and courts
Privilege exposure	Matter material is entered into a service outside the firm's contractual perimeter	Rarely, and usually late
Coverage illusion	The tool reports accurately on the documents supplied, and the population is never tested	No
Fluency deference	A generated draft is reviewed less rigorously than a human draft of the same quality	No
Barrier crossing	Retrieval surfaces material from a matter the person is walled from	No
Skills erosion	Routine delegation removes the work through which junior lawyers learn to reason	No, for several years

TABLE 3 The right-hand column is the point of the table. A failure that produces a document looking exactly like good work will pass supervision, so it has to be prevented by design.

4.1 Fabricated authority, and why the usual remedy fails

Every firm has responded to this risk with an instruction to check. Instructions to check do not survive deadline pressure, which is why the published cases involve competent practitioners.

The control that works is structural. Verification becomes a step in the workflow with a name against it, performed against a citator, never against the tool that produced the citation, and recorded on the file. A document that reaches a court passes through a gate where somebody has confirmed that every named authority exists, is current, is correctly cited and says what the submission claims it says. **ECAVEO**

That is more than an instruction and less than a technology project. It is a checklist with a signature, and the reason it works is that it produces a record which makes the absence of verification visible.

4.2 The scale of the problem, stated carefully

The pattern was established early. In the United States a federal judge fined two attorneys and their firm \$5,000 in June 2023 after a brief cited airline cases that did not exist, with fabricated excerpts supplied when the citations were challenged.¹⁹ **CASE LAW**

A publicly maintained tracker of decisions involving fabricated citations recorded 2,036 cases worldwide as at 10 September 2026, of which the United States accounted for 1,393 and the United Kingdom for 69, with rather more cases involving self-represented litigants than lawyers.¹² **SURVEY**

Three qualifications matter. The compiler states that the collection seeks to be exhaustive while remaining a work in progress that will expand as new examples emerge. Some entries reflect a judgement about unconfirmed allegations, short of a definitive judicial finding. The count rises as awareness rises, so the trend line partly measures detection.¹² **EVIDENCE GAP** Read as an order of magnitude it is still the best public evidence available. The tracker holds 2,036 cases, 69 of them here.

4.3 Fluency deference

The fourth failure has no published cases because it produces no artefact. A reviewer reading a polished draft applies less scrutiny than the same reviewer reading a rough one, and a generated draft is always polished. The consequence is a quiet reduction in review intensity across every matter that uses one.

I know of no study measuring this in legal work. **EVIDENCE GAP** The control is cheap enough that I would apply it without one. Mark generated text until it is adopted, and require the reviewer to adopt it in terms.

WHERE EACH FAILURE IS CAUGHT

A At the boundary Contracts, approved systems and matter entitlements stop privilege exposure and barrier crossing before they occur.	B In the output Citations that resolve, stated coverage, marked generated text and explicit refusal expose fabrication and illusion.	C At the gate A named verification step before anything leaves the firm or reaches a court.	D In the programme Deliberate apprenticeship tasks and reasoning review address skills erosion, which nothing else will.
---	--	---	--

FIGURE 2 Each failure has one natural interception point. Firms concentrate on the second column and under-invest in the third, which is the one the published cases turn on.

Ecaveo framework. **ECAVEO**

4.4 Skills erosion, and what can honestly be said about it

The concern is that automating the work through which junior lawyers learn to reason removes the apprenticeship, and that a cohort trained on review of machine output will reach seniority without the judgement the role requires. The concern is widely held across the profession and is actively debated in the trade and professional press.

It remains an argument. I could locate no longitudinal competency evidence establishing the effect, and any firm claiming to know the answer is guessing. **EVIDENCE GAP** **PRACTICE** The tribunal decision discussed in Chapter 2 makes the supervisory dimension concrete, observing that a supervisor who fails to prevent a junior's false citations is likely to be more culpable than one who misses their own errors.¹⁰ **CASE LAW**

What a firm can do is deliberate and measurable within one intake. Assign reasoning tasks that a tool cannot perform, review the reasoning itself, and assess whether the capability is forming. That costs supervisor time and produces evidence, which is more than the debate currently has. Whether any firm will fund it during a bad quarter is another question.

CHAPTER FIVE

Privilege, and the control model

The distinction between an open service and a contracted one is now the most consequential technical decision a firm makes.

A tribunal has held that uploading confidential documents into an open tool places the information in the public domain and waives privilege, while distinguishing closed and enterprise systems.¹⁰ **CASE LAW** The regulator has warned that information entered into these systems may be stored, retained or used to improve the tool.⁵ **REGULATOR** The representative body has warned that providers may see input and output data and that data often leaves the United Kingdom.¹⁷ **REGULATOR**

Those three statements point the same way and they turn a policy question into an architecture question. A firm should be able to demonstrate technically, and not merely assert, that matter material cannot reach a service the firm has not contracted with. My own view, on which a firm should take its own advice, is that a written policy without a technical control is unlikely to be treated as adequate once a breach has occurred.

RISK TIERS AND MINIMUM CONTROLS

TIER	TYPICAL SCOPE	MINIMUM CONTROLS
One	Public or non-confidential internal material, with reversible assistance	Approved tool, training, source checking and human review before anything is published
Two	Confidential matter information, or decisions internal to a workflow	Registered use case, data-flow review, evaluation, logging and a named reviewer
Three	Client work product, advice or a transactional document	Matter-level entitlements, contractual controls, representative testing and a named supervising lawyer
Four	Filed documents, formal advice, court submissions and autonomous material action	A named accountable lawyer; independent verification of every authority, strict entitlements, monitoring and a tested rollback

TABLE 4 Tier is set by what the output touches and by what would follow from an error. The sophistication of the underlying system does not enter the assessment.

5.1 The six technical controls

Six controls carry most of the assurance burden, and they are consistent with the security guidelines co-sealed by agencies from 18 countries, the American standards body's generative risk profile and data protection accountability expectations.^{1,2,3} **STANDARD**

1. Enterprise identity and matter-level entitlements are enforced before retrieval and before any tool call.
2. Contracts define training use, retention, hosting region, subprocessors, deletion, incident notification and audit rights, and the absence of any one of them is treated as a finding.
3. Agents operate with least privilege and short-lived credentials, and any consequential action requires explicit approval and can be reversed.
4. Every production system records the model, prompt, retrieval corpus, tools, permissions and evaluation version behind each output.
5. Prompt injection, hostile documents, sensitive-data disclosure and unsafe output handling are tested before scale, using material of the kind the system will meet.
6. Generated work passes the same professional review and records controls as work produced without it.

Two further reference points are worth holding. A firm wanting a certifiable management system has ISO/IEC 42001, published in December 2023.²⁰ **STANDARD** Revised provisions on automated decision-making under the Data (Use and Access) Act 2025 came into force in February 2026 and reach any matter where a decision about a person is taken without meaningful human involvement.²¹ **STANDARD** The government's own playbook sets out ten principles which, although written for departments, work as a check on a firm's own policy.⁴ **STANDARD**

5.2 Conflicts and information barriers stay deterministic

Conflict checking and ethical walls are deterministic controls, and a retrieval system is probabilistic. Placing the second in charge of the first is the error to avoid. A model asked to find relevant precedents will find relevant precedents, including in matters the person is walled from, unless entitlements are resolved before the index is queried.

Systems that retrieve broadly and then suppress results are common and cheaper to build. They are unsuitable for a firm with information barriers, and my own view is that a firm should treat retrieve-then-filter architecture as a disqualifying finding in procurement.

ILLUSTRATIVE SCENARIO

The tool that was approved for one thing

An invented commercial firm approves an enterprise assistant for internal knowledge search. Within six weeks fee earners are using it to draft client correspondence, because it is the only approved tool on the desktop and it is good at drafting.

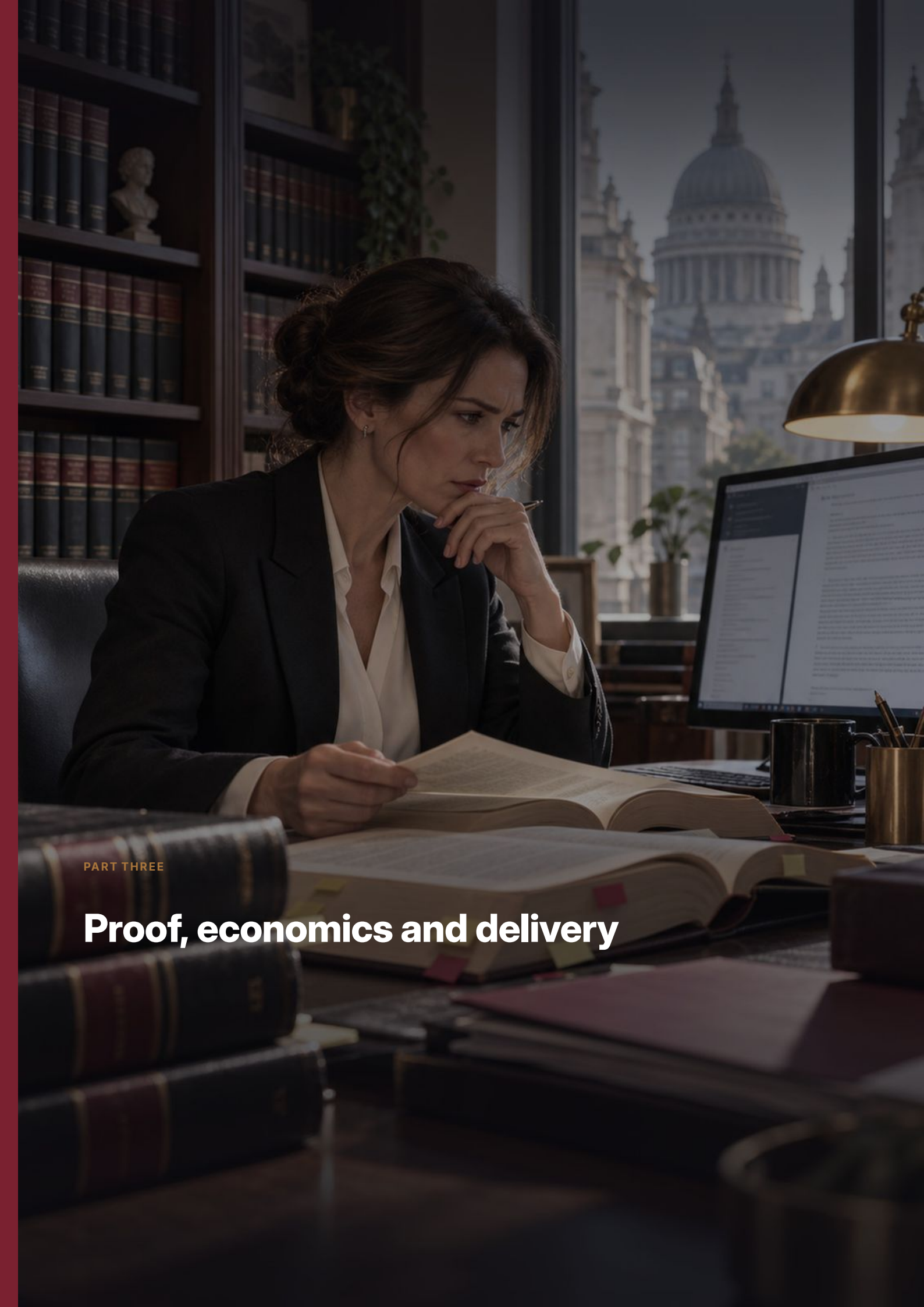
Nothing was breached. The tool is contracted, the tenancy is correct and the entitlements hold. What failed is the certification, which covered knowledge search and said nothing about client-facing output, so no reviewer knows that the correspondence passing their desk was drafted by a system nobody tested for that purpose.

Approval for a purpose is not approval for a tool, and a firm that records only the second has no way to notice the difference.

5.3 What to ask a supplier, in the order that matters

Four questions do most of the work in procurement. What happens to matter material, and where in the contract does it say so. How is the system's behaviour versioned, and how are we told when it changes. What does the system do when the corpus does not support an answer, and can a user see that state. Who at your firm is accountable when the output is wrong.

I ask the third first. A system that cannot express uncertainty will not be safe on a tier three matter whatever else it does well, and the reliability evidence in Chapter 1 is the reason.

A woman with dark hair tied back, wearing a dark blazer over a white shirt, sits at a desk in a law firm. She is looking down at an open book, with her hand resting on her chin in a thoughtful pose. The desk is cluttered with several other open books, some with yellow sticky notes. To her right, a computer monitor displays a legal document, and a brass desk lamp provides warm light. In the background, a large window offers a view of a city skyline, featuring a prominent domed building. To the left, a bookshelf is filled with numerous legal volumes, and a small bust is visible on one of the shelves.

PART THREE

Proof, economics and delivery

CHAPTER SIX

Proving it, and the evidence gate

Write the protocol before anybody sees the tool, and have the review done by a lawyer who does not know which output came from where.

THE SIX DIMENSIONS OF AN EVIDENCE GATE

DIMENSION	WHAT IS MEASURED	THE GATE
Outcome	Cycle time, coverage, search time or throughput	Improves the intended workflow against a real baseline
Quality	Authority validity, currency, jurisdiction, completeness and rework	No material regression, with fabricated or unsupported authority at zero
Judgement	Challenge quality, override behaviour, escalation and reviewer understanding	Lawyers retain independent judgement and can explain the result
Confidentiality	Permission leakage, matter-boundary tests, retention behaviour and red-team findings	No unresolved critical finding, with residual risk accepted by a named owner
Economics	Licence, compute, integration, assurance, review and support cost	Total cost is justified by verified benefit, including the cost of checking
Operations	Reliability, supplier changes, monitoring, support and rollback	A named owner and tested operational controls are in place

TABLE 5 A pilot passes when every dimension clears its gate. Partial passes are how firms scale something that suits the team that chose it.

6.1 Three pilots worth running

An authority-verified research assistant, scoped to answering a narrow research question from an approved primary-source corpus, tested on authority validity, currency, jurisdiction, completeness and the rate of appropriate refusal.

A contract issue spotter, scoped to identifying and citing defined clauses in a permissioned agreement set, tested on clause recall, false positives, review time and confidentiality behaviour under deliberate probing.

A matter chronology builder, scoped to producing a dated, source-linked chronology from a controlled bundle, tested on event recall and date accuracy, on source linkage and stated coverage, and on how much correcting a lawyer has to do.

The quality gate for the first of these should be zero on fabricated authority. Any pilot that produces one has failed, and I would not treat it as needing adjustment. **ECAVEO** That is a stricter standard than the other papers in this series apply, and Chapter 2 is the reason.

6.2 The minimum assurance record

Six things, for every production system. Purpose, intended users, benefit, owner and review date. Allowed and prohibited uses, data classes, jurisdictions and actions. Model, supplier, prompt, retrieval configuration, connectors, permissions and dependencies. Risk assessments, contractual controls and professional decisions taken. Evaluation sample, method, results, limitations and approval. Monitoring, feedback, incidents, changes, rollback and retirement.

6.3 Scale, redesign or stop

Scale where the benefit is achieved, quality is maintained, residual risk has been accepted by a named owner and ownership is funded. Redesign where a value signal exists and quality, review burden, workflow fit or controls miss the gate. Stop where no verified benefit appears, the risk is unacceptable, the fit is weak, or a simpler method performs better.

Ask how many pilots the firm has stopped. The answer is usually none.



Multidisciplinary oversight. The people who decide what a tool may be used for should include the people who will answer for it.

CHAPTER SEVEN

Verification, economics and the fee

Two problems sit here. Verification is expensive and nobody budgets it, and efficiency in an hourly market raises a question the profession here has not answered.

7.1 The economics of verification

A business case that counts hours saved in production and stops there is counting the wrong side of the ledger. Generation becomes nearly free, and the binding constraint moves to verification, performed by the senior lawyers whose time is the most expensive in the firm. **ECAVEO**

The test applies to any proposal in about a minute. Take the cost of producing the work by hand, the cost of generating it, and the cost of checking the generated version to the standard the work requires. Where generation plus verification comes in below production, the case holds.

WHY SOME USE CASES PAY AND OTHERS DO NOT

1 Cheap to check Extraction with a pointer to the clause, or research with a citation that resolves. A lawyer opens three and knows.	2 Costly to check An unsourced summary of a bundle. Checking it means reading the bundle, which is the work it replaced.	3 Impossible to check An answer on the law with no cited authority. There is nothing to verify against, and the reliability evidence says it will sometimes be wrong.
---	---	--

FIGURE 3 Verification cost is a design property. A resolving citation moves a use case from the second column into the first.

Ecaveo framework. **ECAVEO**

Read that way, the citation stops being a courtesy. It is the mechanism that makes a use case economic, because it converts verification from re-performance into spot-checking.

7.2 The fee question

A professional ethics opinion issued in the United States in July 2024 addresses the point squarely. It states that a lawyer may not bill a client for time saved through these tools while charging as though the time had been spent, that a fee charged for which little or no work was performed is an unreasonable fee, that subscription costs must be treated either as overhead or as a disclosed client expense, and that a lawyer may not charge a client for their own learning curve.¹¹ **ADVISORY** A state bar opinion of January 2024 reaches compatible conclusions and adds that a lawyer must tell the client, preferably in writing, of an intention to charge the actual cost of using these tools.¹⁸ **ADVISORY**

No equivalent guidance exists in England and Wales so far as I could establish. **EVIDENCE GAP** That absence is itself a planning point. A firm operating on hourly rates that materially reduces the time a task takes has a conversation coming with its clients, and the firms that open it will do better than the firms that wait to have it opened for them.

The position has been criticised, and the criticism is worth understanding. Commentators have argued that tying reasonableness to time entrenches time-based billing at precisely the moment the profession claims to be moving towards value, and that it creates an incentive against adoption.¹¹ **PRACTICE** That is professional opinion, and a firm should treat it as a live argument.

7.3 The cost lines that are missing

Three costs are routinely absent from a pilot case. Senior review time, described above. Assurance work, meaning evaluation, red-teaming and supplier diligence, with the records supporting them. The cost of supplier change, meaning the regression testing and re-approval that follows every material model update, which recurs indefinitely.

My own rule of thumb, offered as practitioner judgement rather than measured fact, is that a tier three deployment carries ongoing assurance and review costs of a similar order to its licence cost. **PRACTICE**

EVIDENCE GAP

I found no statement from the regulator or the representative body in England and Wales addressing charging for work produced with these tools. The debate here is currently American.

CHAPTER EIGHT

Ninety days, then nine months

The first month is observation and inventory. A programme that starts by choosing a vendor produces a tool nobody is permitted to use on a matter.

THE DELIVERY SEQUENCE

PHASE	TIMING	OUTPUTS
Discover	Weeks 1 to 4	Matter workflows, client terms, information barriers, current tools and training needs mapped, with an inventory of current use
Prove	Weeks 5 to 10	Three narrow pilots on approved sources and representative matter tasks, with blinded review and written stop conditions
Standardise	Weeks 11 to 18	Matter-level controls, verification rules, tool register, supplier terms and supervision guidance
Scale	Months 5 to 9	Expansion by practice group, with client communication, pricing policy and measures on skills formation

TABLE 6 The Standardise phase is where verification rules, matter-level controls and supervision guidance become the default.

8.1 The first 30 days

Interview practice leaders and observe real work across intake, research, review, drafting and litigation preparation. Inventory what is already in use, including tools embedded in software the firm already licences, and including the unsanctioned use that every survey in this field suggests is present. Map matter boundaries, client terms, entitlements and the lawyers who actually decide.

Publish interim guidance on safe use within the first fortnight, together with a simple route for proposing an experiment, because a shadow estate grows while a programme designs itself. The interim guidance should do one thing above all others, which is to state plainly which services may receive matter material and which may not.

8.2 What success looks like at nine months

Scaled workflows with measured quality, time or service improvement against the baselines set in month one. A single intake route, approved systems, documented use cases and monitored supplier changes. Lawyers who can explain what a tool cannot do, verify its output and escalate a failure. A verification gate that leaves a record on every matter that reaches a court. A backlog separating proven opportunities from research and from uses the firm has decided against.

CHAPTER NINE

The session to run with the partners

Eight questions, applied to any proposed use. A use case that survives all eight is worth funding.

1. What observable outcome should improve, for whom, and by when?
2. What is the current baseline, and where does the work wait, repeat or fail today?
3. What data is involved, and what client and professional obligations apply to it?
4. What error would matter, who would be harmed, and how would anyone detect it?
5. Who reviews the output, who decides, who can override, and who owns the advice?
6. What simpler method should this be compared against?
7. How will a change in the model or the supplier be detected during a matter?
8. What evidence would justify scaling this, and what evidence would stop it?

The fourth and the eighth do most of the filtering. A proposal whose sponsor cannot describe the error that would matter has not been thought through, and a proposal with no stop condition has already committed the firm.

TWO THINGS TO DECIDE IN THE ROOM

Name the lawyer accountable for the verification gate, meaning the step at which every authority in a document leaving the firm is confirmed to exist and to say what is claimed. Then decide, in writing, which services may receive matter material. A tribunal has already held that an open tool is the public domain.¹⁰ A firm that has not drawn that line is relying on the judgement of every fee earner, every day.

CHAPTER TEN

Where the evidence is thin

Four things here would change if the evidence changed, and the weakest link in the argument is the authority I have relied on most heavily for privilege.

10.1 The privilege point binds below the High Court

The clearest English authority on privilege and open tools is a reported decision of the Upper Tribunal.¹⁰ **CASE LAW** That is a superior court of record whose reported decisions bind the First-tier Tribunal, and it is not binding on the High Court or the Court of Appeal. A higher court could reason differently, particularly on the distinction between an open service and a contracted one.

For practical purposes the distinction makes little difference, because a firm should not be putting privileged material into an open consumer service whatever the authority says. It matters for how the point should be described in a policy document, and it is the clearest example in this paper of a place where a firm should take its own advice on the point.

10.2 The reliability figures are contested

The measured hallucination rates in Chapter 1 come from a single research group, and both vendors disputed the original publication.¹³ **VENDOR RESEARCH** One objection has been answered, since the study was extended to the disputed product and peer-reviewed. The other, which concerns how broadly hallucination is defined, has not been resolved and reasonable people can differ on it.

A materially better result on a later generation of these products would change the procurement argument and would leave the workflow argument intact. No plausible result removes the verification step, because the published cases involve tools performing as designed on questions they were not equipped to answer.

10.3 Adoption figures are nearly all vendor-produced

Every headline adoption statistic I could locate for 2025 and 2026 comes from a supplier to the legal market, and none published a sampling frame or response rate.^{14,15,16} **VENDOR RESEARCH** They are directional. The one figure that cuts against its publisher's commercial interest, showing legal-specific tool use falling while overall use held steady, is the more credible for that reason and deserves a firm's attention.

10.4 The two questions nobody has answered

Whether these tools erode the reasoning capability of junior lawyers is unresolved, and I have found no longitudinal evidence either way. **EVIDENCE GAP** Whether broader coverage in document review produces better outcomes, or simply more findings to discount, is equally unresolved. **EVIDENCE GAP**

Both are testable inside a single firm within two intakes, and a firm that keeps a record of what it decided and why will be able to answer them for itself long before the profession answers them collectively. That is the strongest argument in this paper for starting the record now, and it is the argument that runs through all three papers in this series.



Client service. Every control in this paper exists to protect the moment at which a client is told what to do.

References

Guidance and case law in this area remain subject to jurisdiction, to the facts of a matter and to later development. Where a source is a survey run by a firm with a commercial interest in the result, the reference says so.

1. Information Commissioner's Office. *Guidance on AI and data protection*. Page last updated 15 March 2023 and flagged by the Commissioner as under review. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/>
2. National Cyber Security Centre. *Guidelines for secure AI system development*. 27 November 2023, co-sealed by agencies from 18 countries. <https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development>
3. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1. 26 July 2024. <https://doi.org/10.6028/NIST.A.1.600-1>
4. UK Government. *Artificial Intelligence Playbook for the UK Government*. 10 February 2025. <https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government.html>
5. Solicitors Regulation Authority. *Misuse of AI* warning notice. 17 August 2026. <https://www.sra.org.uk/solicitors/guidance/misuse-ai/>
6. Solicitors Regulation Authority. *Risk Outlook report: The use of artificial intelligence in the legal market*. 20 November 2023. <https://media.sra.org.uk/sra/research-publications/artificial-intelligence-legal-market/>
7. Courts and Tribunals Judiciary. *Artificial Intelligence (AI): Judicial Guidance*. October 2025, replacing versions issued in December 2023 and April 2025. <https://www.judiciary.uk/guidance-and-resources/artificial-intelligence-ai-judicial-guidance-october-2025/>
8. Civil Justice Council. *Use of AI for Preparing Court Documents: Interim Report and Consultation*. February 2026, with a consultation update published in July 2026. Final report awaited as at September 2026.
9. *Ayinde v London Borough of Haringey; Al-Haroun v Qatar National Bank* [2025] EWHC 1383 (Admin). Divisional Court, 6 June 2025. <https://www.judiciary.uk/wp-content/uploads/2025/06/Ayinde-v-London-Borough-of-Haringey-and-Al-Haroun-v-Qatar-National-Bank.pdf>
10. Upper Tribunal (Immigration and Asylum Chamber), [2026] UKUT 00081 (IAC). Promulgated 17 November 2025. <https://tribunalsdecisions.service.gov.uk/utiac/2026-ukut-00081>
11. American Bar Association. *Formal Opinion 512: Generative Artificial Intelligence Tools*. 29 July 2024. Criticism of the fees analysis has been published by professional commentators and is opinion rather than authority.
12. Charlotin, D. *AI Hallucination Cases*. Public database, count of 2,036 cases as at 10 September 2026. Described by the compiler as seeking to be exhaustive while remaining a work in progress. <https://www.damiencharlotin.com/hallucinations/>
13. Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., and Ho, D. E. *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*. arXiv:2405.20362, 30 May 2024, subsequently published in the Journal of Empirical Legal Studies. Both vendors publicly disputed the findings.
14. LexisNexis. Survey of UK legal professionals, conducted January 2025. Vendor survey with undisclosed sampling frame and response rate.
15. LexisNexis. Survey of 543 UK legal professionals, reported September 2026. Vendor survey with undisclosed field dates and method.
16. Clio. *2025 Legal Trends Report*. 16 October 2025. United States sample. Vendor survey, reported here for the finding that legal-specific tool use fell while overall use held steady.
17. The Law Society. *Generative AI: the essentials*. Page dated September 2025. <https://www.lawsociety.org.uk/topics/ai-and-lawtech/generative-ai-the-essentials>
18. The Florida Bar. *Ethics Opinion 24-1*. 19 January 2024. Advisory opinion. <https://www-media.florida-bar.org/uploads/2024/01/FL-Bar-Ethics-Op-24-1.pdf>
19. *Mata v Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. 2023). Sanctions order of 22 June 2023, imposing a \$5,000 fine.
20. International Organization for Standardization and International Electrotechnical Commission. *ISO/IEC 42001:2023, Artificial intelligence management system*. December 2023.
21. United Kingdom. *Data (Use and Access) Act 2025*. Provisions on automated decision-making commenced 5 February 2026.

ECAVEO RESPONSIBLE AI SERIES

Verified Authority

Paper three of three. Version 1.0, September 2026.

ALSO IN THIS SERIES

Paper one. The Traceable Thesis. Artificial intelligence across the private equity investment lifecycle.

Paper two. Sufficient and Appropriate. Artificial intelligence, audit evidence and professional scepticism.

THE AUTHOR

Paul Forrest is a fractional Head of AI working with private equity and venture capital portfolio businesses on the deployment of artificial intelligence.