ClinLexis™
READ · ANALYSE · REPORT

WHITEPAPER · MEDICO-LEGAL PRACTICE

Reviewing medical records in the age of Artificial Intelligence

The hidden bottleneck in medico-legal work, the shadow AI shortcut that is now in routine use, and what a responsible alternative actually looks like, for solicitors and expert witnesses.

AUTHORS

Dr Khaled Abdel-Aziz, PhD, FRCP · Consultant Neurologist & Medical Co-founder, ClinLexis

Dr Ahmmad Youssef, PhD · Technical Co-founder, ClinLexis

PUBLISHED

May 2026 · ClinLexis Ltd · United Kingdom

Executive summary

Three things every medico-legal firm now has to take a view on: a bottleneck consuming expert and solicitor time on every serious case; an unauthorised AI shortcut already in use; and the evidence that distinguishes a responsible tool from one that quietly transfers risk back to the user.

1. **Medical-record review is the hidden bottleneck in modern medico-legal practice.** NHS documentation volumes have risen sharply over the past decade, and a typical medical-negligence or serious personal-injury case requires an expert to absorb several thousand pages before any opinion can be formed. The work that delays cases is not the formation of expert opinion. It is the manual extraction and organisation of facts that has to happen first.

2,000–4,000

pages routinely required for an expert opinion in a serious medical-negligence or personal-injury case; bundles over 12,000 pages are no longer remarkable.

2. **A consumer-chatbot shortcut is in use across both branches of the profession, with the scale hard to measure.** ChatGPT, Claude, Gemini and Copilot, on their free or personal-paid tiers, were not designed for the special-category data being pasted into them. Use of these tools for client work creates confidentiality, data-protection and professional-duty exposures that procurement processes have not caught up with.

€15m

fine imposed by the Italian data-protection authority on OpenAI in December 2024 for handling of personal data in ChatGPT.

3. **A responsible alternative exists, and the difference is in the engineering rather than the marketing.** An AI system fit for medico-legal work is one whose claims about a record can be verified at the source in one click, whose reasoning is visible in plain English, and which stops short of clinical or legal opinion. These properties are observable from the outside, and any vendor in this space should be able to describe them in writing.

THE CONCLUSION

The question for solicitors and instructing partners is no longer whether their teams (and the experts they instruct) are using AI. In all likelihood, they are. The question is **whose AI, on what terms, and with what evidence**. Responsible AI in this domain is engineered, not asserted: every claim about a record traces back to the page that produced it in one click, the reasoning is visible in plain English, and the system stops short of clinical or legal opinion.

1. The bottleneck behind every medico-legal case

An expert instructed in a medical-negligence or personal-injury case cannot form an opinion until they have understood the chronology: what happened, when, and how those events relate to one another. That is the foundation on which any view on breach, causation or prognosis is built.

In a straightforward case, this can be quick. In the cases that actually drive the work of a serious medico-legal firm (catastrophic injuries, complex pre-existing comorbidity, fragmented care across multiple Trusts) it is anything but. Bundles of two to four thousand pages are routine. Bundles in excess of twelve thousand pages, while rare, are no longer remarkable.

Even with the move from boxes of lever-arch files to electronic bundles, the underlying task has not changed. The information is fragmented, sometimes repetitive, occasionally incomplete, and regularly formatted in ways that demand careful interpretation. Reading the material is only the start. Before an opinion can be formed, the expert has to:

- read through large volumes of unstructured material;
- identify the relevant clinical events;
- establish a reliable timeline;
- cross-reference entries across different sources; and
- distinguish between material and non-material information.

Only then can the meaningful analysis begin: the work the expert was instructed for.

For an instructing solicitor, the consequence of this is well known but rarely named directly. The principal source of delay in medico-legal work is not the writing of expert reports. It is the time the firm itself spends reviewing the records and assessing the case before an expert is even instructed, and the time the expert then spends building a factual chronology before a single sentence of the report can be written. That preparation cascades into everything else: the date a claim can be assessed for merit, the date a Part 36 offer can be made or responded to, the budget headroom remaining when the matter reaches a Costs and Case Management Conference (CCMC), the date a settlement meeting can be convened. **A bottleneck at the front of the process is felt at every milestone after it.**

Capacity is itself a constraint. The strongest experts have long waiting lists, bounded by their NHS clinic schedules. Reducing the preparation time per case is one of the few levers that shortens those lists.

The volume problem is structural and getting worse. NHS clinicians produce more documentation per encounter than a decade ago, and patients move between providers more often, so records are dispersed across multiple systems. Instructing firms face tighter cost-budget discipline, faster turnaround expectations, and increasingly complex cases. Workflows that depend on highly skilled professionals manually extracting and organising information are not sustainable.

Other areas of law have already faced this. Electronic disclosure transformed how large document sets are handled in commercial litigation; the professional judgement at the heart of the work did not change, but the mechanics around it did. Medical-record review is approaching a similar inflection point, and the aim is not to replace clinical or legal judgement but to relieve the mechanical work around it.

2. The shortcut professionals are quietly taking

Given a deadline, a full caseload and a free chatbot one tab over, the natural response to the bottleneck described above is to use the chatbot. This is happening across the profession, even if the scale is hard to measure.

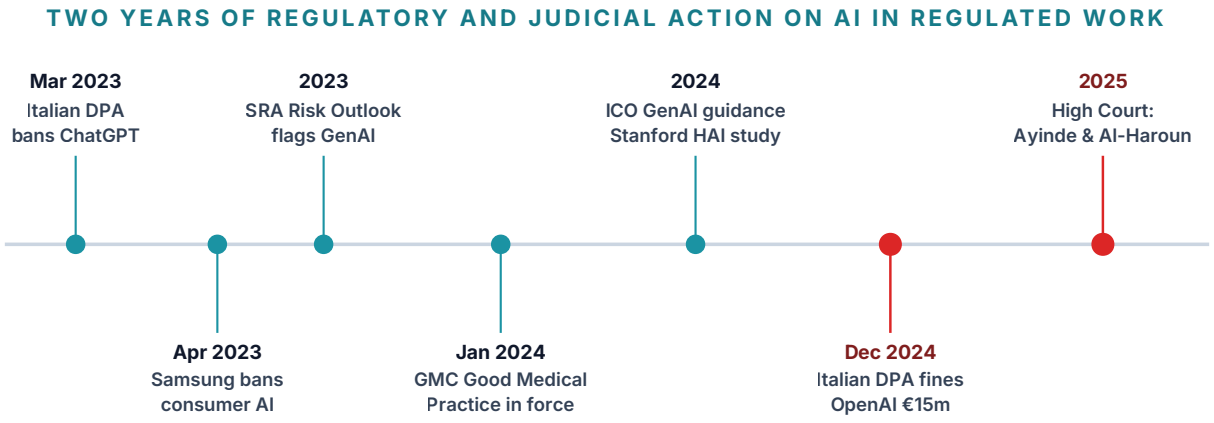
We are not making a moral point. Although the tools are useful, free, instant, and require no IT approval, the architecture beneath consumer chatbots was not designed for the data being pasted in. That gap is now arguably the most underestimated AI risk in medico-legal work, and one the regulators have already drawn a line under.

The evidence that this is not a hypothetical

Regulators have responded. Within the past two years the Solicitors Regulation Authority, the Bar Standards Board and the General Medical Council have each issued guidance specifically on the use of generative AI by their members. The SRA's 2023 Risk Outlook on Technology and Innovation singled out generative AI as a live area of concern and reminded firms of their existing confidentiality and supervision duties. Regulators do not write rules against theoretical problems.

Usage is high across both branches of the profession. Sector surveys through 2024, including the Law Society's own work and the annual reports from the major legal-information publishers, have consistently shown a substantial proportion of UK fee-earners using generative AI tools regularly, with usage particularly high among trainees and junior associates. Equivalent surveys of NHS clinicians over the same period show similar patterns of personal experimentation. These figures cover all generative AI use, not only consumer tools, but the consumer share is unlikely to be small. Consumer chatbots are dramatically easier to reach than vetted enterprise alternatives.

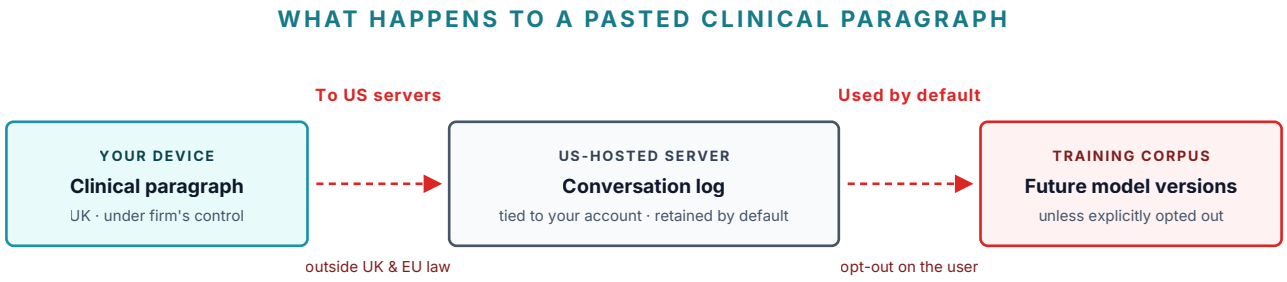
The courts have already seen the consequences. Ayinde and Al-Haroun, both decided in the High Court in 2025, exist because legal professionals used consumer AI tools to draft material that was then placed before the court without verification. Wasted-costs orders and regulatory referrals followed. In the United States, Mata v. Avianca (2023) and several follow-on matters show the same pattern. The cases that became public are easy to count. The cases that did not, by definition, are not.



Where the data actually goes

When a paragraph from a clinical record is pasted into the consumer version of ChatGPT, several

things happen, none of them visible to the person doing the pasting.



The text leaves the device and is sent to OpenAI's servers, which sit in the United States. Under the consumer terms (free and Plus tiers, by default), conversations are retained and may be used to improve future models, unless the user has explicitly opted out. The conversation is associated with the user's account and stored on infrastructure that is not subject to UK data-protection law in the same way a UK-resident system would be. Broadly the same is true of consumer Claude, Gemini, Copilot in its consumer tier, and the various Mistral and Llama-based assistants. Defaults differ, retention policies differ, opt-out mechanisms differ. None of these products were designed with special-category personal data in mind.

The contracted enterprise versions of these tools are different. OpenAI's enterprise products and Microsoft's Azure OpenAI both contractually exclude customer data from training, and the latter offers UK regional deployments. Anthropic's enterprise tier is no-training by default. These are real, careful products. They are also not the products people open on their phones at lunchtime.

The regulators have been clear on the risk. The Italian data-protection authority temporarily banned ChatGPT in March 2023 for breaches of GDPR around lawful basis, transparency and the handling of personal data; in December 2024, the same regulator imposed a €15 million fine on the same grounds. The Information Commissioner's Office (ICO) 2024 guidance on generative AI and data protection cautions specifically against the use of consumer generative AI tools for special-category data. The most public corporate failure remains Samsung in April 2023, where engineers pasted proprietary source code and meeting notes into ChatGPT and the data left the company's control. Samsung banned consumer generative AI for internal work within weeks. JPMorgan Chase, Verizon and several others have taken the same position.

Why this matters more in medico-legal work

For most professions, this is a confidentiality and intellectual-property issue. For medico-legal work, it is something more.

Medical records are special-category personal data under Article 9 of the UK GDPR. Special-category processing requires both a lawful basis under Article 6 and a separate Article 9 condition; for litigation, usually 9(2)(f), the establishment, exercise or defence of legal claims. None of those conditions relax the underlying obligations. Encryption, access control, retention limits and a documented Data Protection Impact Assessment are still required.

When a fee-earner pastes a discharge summary into a consumer chatbot, those obligations do not pause. The processing is taking place. The data is leaving the firm's control. There is no Data Protection Impact Assessment (DPIA). There is no audit trail. The firm cannot tell a regulator what was sent, when, by whom, or what came back. If a patient subject access request later asks where their data

has been processed, the firm cannot honestly answer.

The professional duties run in the same direction. SRA Principle 7 requires solicitors to act in each client's best interests, which the SRA has interpreted to include using only systems that protect client information. The GMC's Good Medical Practice (in force from January 2024) requires doctors to maintain confidentiality and to use only secure systems for patient information. The Bar Standards Board has issued similar guidance. And then there is Ayinde: the High Court has already made plain that **the duty to verify AI output cannot be delegated**. The duty of confidentiality is older, deeper and equally non-delegable. Both run in the same direction. A tool that cannot be inspected, audited or controlled is a tool that cannot lawfully be used for client work.

Why this is a procurement problem, not a training problem

It would be tempting to write the rest of this section as a finger-wag at junior fee-earners or harried experts. We do not think that is useful. The people doing this are responding to a tooling gap, and writing them off does nothing to close it.

The unaided alternative, reading the bundle by hand at 11pm after a full caseload, is a poor experience for everyone involved. The expert who asks a chatbot to plain-English a discharge summary at 7am before clinic is trying to do their job within the time available. That does not make the act compliant with their professional duties; it makes it predictable.

Many firms have an "AI policy" that consists of a paragraph in the staff handbook saying not to use consumer AI for client work. The policy is unenforceable, the supervision is patchy, and the workflow it implicitly demands (back to manual reading) is not realistic. The result is shadow AI use on a scale most managing partners would not want to admit, and would not be able to evidence to a regulator if asked.

> **THE HONEST POSITION**
>
> The profession needs alternatives, not warnings. Professionals will reach for AI; the question is which AI. The job of the firm, and of any vendor in this space, is to make sure the responsible alternative is at least as easy to use as the consumer one.

3. What "responsible AI" actually means in medico-legal work

When people say AI is a "black box", they usually mean something specific: you put a question in at one end, an answer comes out the other, and what happens in the middle is hidden. You are being asked to trust a result without being able to see how it was reached. For a profession whose work depends on being able to defend its reasoning, that is a serious concern, and an entirely reasonable one.

It is not, however, the whole picture. The black-box framing fits a single AI model in isolation. A product built around a model is a different thing: it is a pipeline of separate steps, and each step can either hide its working or show it. Whether those steps are visible or opaque is one of the most important decisions an AI vendor makes. In medico-legal work, that decision is the difference between a tool a solicitor or expert witness can adopt and one they cannot.

Where black-box AI is already not accepted

The principle that AI should be inspectable, not just impressive, is not new and not unique to this profession. Several other domains have already drawn the line.

In clinical care, the Medicines and Healthcare products Regulatory Agency (MHRA) regulates AI as a medical device, and explainability is part of what developers have to demonstrate before deployment. UK GDPR Article 22 reinforces this from the patient's side, giving anyone subject to a solely automated medical decision the right to "meaningful information about the logic involved" and to human intervention. NHS England's guidance for digital and data-driven health technology requires AI systems used in NHS care to be transparent in operation. When a radiology tool flags a region on a mammogram, the clinician sees the highlight, the confidence score and the visual overlay, not just a verdict.

Financial services drew the same line earlier. The Financial Conduct Authority's (FCA) Consumer Duty obliges firms to support customer understanding of decisions affecting them, and the Bank of England's supervisory statement on model risk management (SS1/23, in force from May 2024) requires firms to evidence the interpretability of their AI and machine-learning models. Regulators were never going to accept "the algorithm declined the loan" as a final answer.

The newest large language models themselves have moved on. Several now expose their intermediate reasoning rather than only the final answer, and national AI safety institutes (including the UK's AI Security Institute) are publishing evaluations of how faithfully those traces reflect what the model actually does. The direction of travel across the industry is towards systems whose reasoning can be inspected.

Medico-legal practice should expect the same standard. The verification duty is explicit; the data is the most sensitive there is; and the consequences of being wrong end up in court.

The 2024 Stanford finding, and why it matters

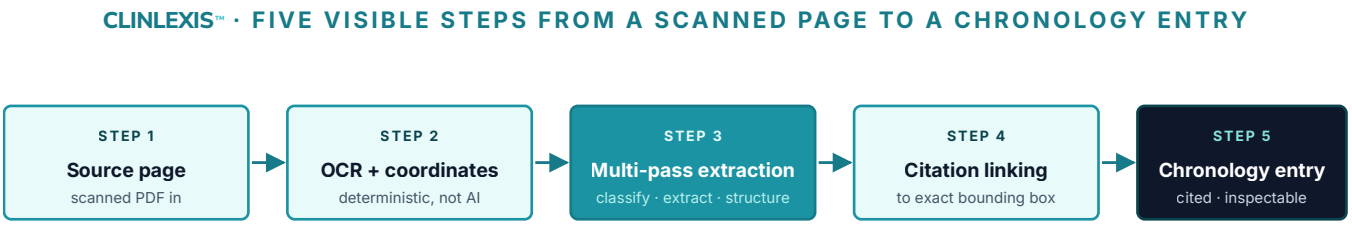
The Stanford HAI study by Magesh and colleagues (2024) tested the leading AI legal-research tools, including ones marketed as "hallucination-free", and found error rates of between **17% and 33%** on retrieval-augmented systems built on curated case-law databases. Lexis+ AI was at the better end; Westlaw's AI-Assisted Research at the worse. These were paid, professional-grade tools sold to qualified solicitors. Models have improved meaningfully since, but the architectural lesson generalises: the vendors were not lying about the underlying architecture; they were over-promising about what it could guarantee. The only durable defence, at any error rate, is a system in which every claim can be checked against the source in one click.

Take that finding alongside Ayinde, and the design objective for any serious AI tool in this domain becomes clear. The duty to verify cannot be delegated to a machine. So the system has to make verification fast, obvious and unavoidable. Every claim the system makes about a clinical record has to be one click from the page that produced it. If that click is hard, or imprecise, or absent, the system has done a bad job, no matter how clever the underlying model.

One fact, end to end: how ClinLexis™ handles it

To make this concrete, here is how ClinLexis handles a single clinical fact end to end. Imagine the document in front of you is a 47-page hospital discharge summary, scanned at a GP surgery, slightly skewed. Buried on page 19 is a sentence: "Lithium level taken 12/03/2024, result 0.4 mmol/L, no

review documented." An expert needs to know about this entry. The five architectural decisions below determine whether the system can be trusted with it — and they are also a useful checklist for evaluating any AI tool in this space.



Each step is independently inspectable. Failure modes are observable, not mysterious.

Optical character recognition with coordinates. Before any language model is involved, the page is processed by optical character recognition (OCR) that returns, for every word, the coordinates of its bounding box. The text is not "extracted by the AI"; it is read off the pixels by a deterministic process that records exactly where each word sat. Those coordinates travel with the text through the rest of the pipeline. They are how the system will find the words again later.

A multi-pass pipeline, not a single model. The popular image of generative AI is one giant model being asked one giant question. That is not how a careful system works. Inside a properly designed product, a single page is processed by several passes, each with a defined input and output: a classification pass that decides what kind of document this is; an extraction pass that picks out clinical events and their dates; a structuring pass that places them on the timeline; a linking pass that matches each event back to its source coordinates; and where appropriate, a reasoning pass that flags possible significance. Each pass can be inspected, replayed and disagreed with. None of them is asked to "do everything". It is unglamorous, multi-stage engineering. It is also why the failure modes of such a system are observable rather than mysterious.

A citation that opens the page, not just the document. When the lithium-level entry appears in the chronology, the citation next to it should not be an aspiration. It should be a link to the exact bounding box on page 19 of the discharge summary, opening the PDF with the sentence highlighted. Not the page; the sentence. If the highlight does not appear, that is a bug, and a serious one — one that is reported to the team in real time and prioritised for resolution.

A "show your working" view on every entry. Each entry should be accompanied by a plain-English explanation of how it was assembled: which page it was drawn from, what the source text actually says, how the date was resolved. The expert should not have to take any chronology entry on faith.

Inconsistency flags that point both ways. If the system finds that another document in the bundle records a different version of events, it should surface the contradiction with citations to both sources. The user sees the conflict, sees the evidence on each side, and decides which one is right.

That is the whole architecture in one example. Five layers of transparency, each independently verifiable. None of it is magic. All of it is engineering choices a vendor could have skipped.

The line a responsible system will not cross

A responsible system surfaces evidence; **it does not make the call**. It will tell you a review interval was 23 days; it will not tell you that 23 days was negligent. It will tell you that two records say different things about a medication; it will not tell you which one is right. Clinical and legal judgement

carry professional accountability: the expert signs the report, the solicitor signs the pleading, the court holds them responsible. Software that generates those judgements is taking on responsibility it cannot discharge — that is responsibility laundering, not responsible AI. For solicitors, CPR Part 35 places the expert's duty squarely on the expert; software cannot borrow that duty.

4. A framework for choosing an AI tool

Solicitors and instructing partners do not need to learn the engineering to ask useful questions. The following framework consolidates the technical and procurement-level questions a firm should be able to answer before adopting any AI tool that touches client information or patient records.

THREE QUESTIONS ABOUT THE OUTPUT

1. **Can you reach the source (the exact passage on the exact page) in one click?**
   If not, the verification burden the courts have placed on the profession has not been engineered for. It has been pushed back onto the user.
2. **When the system flags something as significant, can you see why, in plain English, with the guideline or the contradictory record cited?**
   If the reasoning is not visible, the system is not a glass box. It is a black box with a friendlier interface.
3. **Where does the system stop?**
   If it offers clinical or legal conclusions, ask who is accountable for them. If it surfaces evidence and stops short of the call, you are looking at a tool that respects what your experts and your fee-earners do.

FIVE QUESTIONS ABOUT THE PLUMBING

1. **Where does the data live?**
   Data should be processed and stored within the UK, in infrastructure subject to UK law, with no transfers to other jurisdictions without explicit safeguards. “Global” is a red flag; “EU-equivalent” is not the same as UK.
2. **Does the data train the model?**
   The vendor should be able to point to a contractual clause confirming that customer data is excluded from any training, fine-tuning or evaluation. Default consumer tiers usually fail this test. Contracted enterprise tiers typically pass it.
3. **Is there an audit trail?**
   The system should keep a record of what was uploaded, what was asked, what was returned, by whom and when, for the moment a regulator, a client or a court asks what happened on a given date.
4. **What does access control actually look like?**
   Multi-factor authentication, role-based access, and the ability to revoke access in one click when a colleague leaves. Unglamorous; also the ones that fail first when something goes wrong.
5. **Will the vendor share its DPIA?**
   The vendor should make its own DPIA available, so the firm's Data Protection Officer (DPO) can rely on it without re-running every question from scratch.

5. Where to draw the line

The line we would draw, if asked, is this.

For genuinely non-sensitive work (a generic question about NICE guidance phrasing, a query about a public legal principle, a draft of an internal email) the consumer tools are fine, with the same caution any cloud service deserves.

For anything that contains client information, patient data, or material covered by professional confidence, **do not paste**. Use a tool that has been bought, contracted for and risk-assessed by your firm. If your firm does not have one, that is a problem to escalate, not to work around.

This is not a counsel of perfection. It is the same line the profession has always drawn between work email and personal email, between firm laptops and home laptops, between the case-management system and a Word document on a USB stick. The tools are new; the duty of confidence the profession has always operated under is not.

For instructing solicitors

Ask your experts what they use. Letters of instruction routinely set out what a court expects of the expert; they are an obvious place to also set out what the firm expects of the tools the expert relies on. A short clause in the standard-form letter of instruction, asking the expert to confirm that any AI assistance used in the preparation of the report meets the firm's confidentiality standards, costs nothing and surfaces a great deal.

Ask your panel firms what their experts use. The same logic applies in reverse. If your firm acts for defendants and instructs panel experts through brokers, the question of what those experts paste into what tool is yours to ask, not theirs to volunteer.

For expert witnesses

The choice of tooling is no longer purely operational. The expert who can name the tool used in preparing a report, point to its architecture, and produce a DPIA on request, is in a stronger position than one who cannot. There is also a capacity dividend: time that no longer goes on manual extraction can go to the work the court is paying for, which means more reports per month and shorter waiting lists.

The practical step is to choose the tool now, document the choice, and have the answer ready in a sentence before someone asks for it.

6. Conclusion

The interesting AI question for the profession is no longer whether to use AI. It is whose AI to use, and on what terms. Two questions follow from this paper, one for each branch of the profession.

For the solicitor. What tool does your firm offer its fee-earners and the experts it instructs as the responsible alternative? If the answer is “we haven't decided”, someone else will decide for you, whether it is a regulator, a court, or a client.

For the expert witness. If an instructing solicitor or a court asked today what AI assistance was used in preparing your last report, what would you write? Have it drafted before the question arrives.

References

- Magesh, V., Surani, F., Narayanan, A., Manning, C., & Ho, D. E. (2024). Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools. Stanford HAI / RegLab.
- Ayinde and Al-Haroun — High Court (Divisional Court), 2025: judicial guidance on solicitors' and counsel's responsibilities for AI-generated material in court filings.
- Mata v. Avianca, Inc., 22-cv-1461 (S.D.N.Y. 2023).
- Solicitors Regulation Authority, Risk Outlook: Technology and Innovation, 2023, and subsequent guidance on generative AI.
- Bar Standards Board, Considerations when using ChatGPT and generative artificial intelligence software based on large language models, 2024.
- General Medical Council, Good Medical Practice (in force from January 2024).
- Information Commissioner's Office, Guidance on generative AI and data protection, 2024.
- UK GDPR, Articles 6, 9 and 22.
- MHRA, Software and AI as a Medical Device Change Programme and associated guidance.
- NHS England, A guide to good practice for digital and data-driven health technologies.
- Financial Conduct Authority, Consumer Duty (PS22/9), in force from July 2023.
- Bank of England / Prudential Regulation Authority, SS1/23 — Model risk management principles for banks, effective May 2024.
- Garante per la protezione dei dati personali, Provvedimento del 30 marzo 2023 and Provvedimento del 19 dicembre 2024.
- Civil Procedure Rules, Part 35 — Experts and Assessors.

ClinLexis™

READ · ANALYSE · REPORT

ABOUT CLINLEXIS

ClinLexis is an AI-assisted platform for medical-record review in medico-legal cases, built for UK solicitors and expert witnesses. It addresses the bottleneck described in this paper directly, while staying within the architectural and procurement standards that responsible practice in this domain requires.

WHERE CLINLEXIS SITS AGAINST THE FRAMEWORK

Chronology and bundle review

Auto-drafted chronologies, structured timelines, and cross-document inconsistency flags. Designed to relieve the manual extraction work that precedes the formation of an opinion, not the opinion itself.

Data handling

Stored and processed within UK and EU jurisdictions, on infrastructure aligned with the standards required for patient records. Customer data is excluded from model training. Multi-factor authentication, role-based access, full audit trail.

Verifiable architecture

Every chronology entry links back to the source passage on the source page. Reasoning is shown in plain English. The system surfaces evidence and contradictions; it does not draft clinical or legal conclusions.

THE AUTHORS

Dr Khaled Abdel-Aziz, PhD, FRCP

MEDICAL CO-FOUNDER

Consultant Neurologist with 22 years' NHS experience, practising at St Peter's Hospital, Chertsey and St George's Hospital, London. Acted as expert witness in over 400 clinical-negligence and personal-injury claims, instructed by both Claimant and Defendant solicitors. Fellow of the Royal College of Physicians; PhD in Neuroscience (UCL); MBChB (Leeds). Background in applied AI research in epilepsy and headache medicine.

Dr Ahmmad Youssef, PhD

TECHNICAL CO-FOUNDER

Two decades leading AI and engineering technology in secure, large-scale environments at major UK financial institutions including JPMorgan Chase, Deutsche Bank and Barclays. PhD in Data Analytics and Software Quality from Brunel University; BSc and MSc in Computer Science from Queen Mary, University of London. Has advised FTSE companies on the responsible adoption of generative AI.

CONFERENCE EXCLUSIVE

Try it on your next case

Scan the code to claim a conference-only trial.

clinlexis.co.uk/conference