Protecting Legal Professional Privilege in the Age of AI

MARCH 20, 2026

Artificial intelligence (AI) tools are now embedded in investigations and compliance workflows and play an increasing role in document review, witness preparation, legal advice drafting and meeting transcription. The speed at which investigations professionals have adopted AI has left little time to consider the effect using these tools may have on the protections afforded by legal professional privilege (LPP) under English and Welsh law.

The introduction of new technology does not alter the fundamental principles of LPP, but it does introduce new risks and means by which privilege may be lost. This article identifies the main aspects of an investigation that may be assisted by AI, the risks introduced by these new capabilities and practical steps that investigations professionals can take to mitigate those risks.

The legal framework

By way of brief reminder, LPP in England and Wales has two principal limbs:

1. **Legal advice privilege (LAP)** protects confidential communications between lawyer and client for the dominant purpose of giving or receiving legal advice. In a corporate setting, 'client' is construed narrowly. Privilege does not extend to communications with all employees, only those expressly authorised to seek and receive legal advice on the company's behalf.[1]
2. **Litigation privilege (LP)** protects confidential communications between lawyer, client and third party where the dominant purpose of the communication is conducting litigation that is reasonably anticipated or already in train.[2]

Central to both heads of LPP are the requirements of confidentiality and dominant legal purpose. Both are placed under direct strain by AI tools that encourage the potential non-confidential sharing of information with technology providers and that risk obscuring the dominant purpose of an activity through its automation.

Confidentiality and waiver

Disclosure to a third party, whether deliberate or inadvertent, risks destroying the necessary quality of confidence in a communication and with it any claim to privilege. Every point at which an AI tool processes confidential privileged information is therefore a potential point of failure. The fact that the relevant act is performed by software rather than a person does not negate the legal risk; it simply makes it harder to see and harder to unwind.

The sections below focus on the AI touchpoints that arise most often in internal investigations and disclosure exercises and recommends controls that best preserve confidentiality and the dominant legal purpose.

Document review and e-disclosure

Technology-assisted review, predictive coding and large language model review tools are now common in large disclosure exercises and internal investigations. Using AI to review documents does not, by itself, change the privilege status of the underlying documents. Rather, it enhances the existing risks of inadvertent production, third-party disclosure and the creation of discoverable derivative materials.

1. **Inadvertent production.** As with human reviewers, AI review tools are fallible and prone to misclassifying documents. A model trained to identify documents for production will inevitably mislabel privileged documents as non-privileged, especially where privilege turns on context and nuance (e.g. mixed commercial/legal emails, in-house counsel threads or legal advice embedded in business communications).

As a control, review teams should ensure a human validation step before any production set is finalised and implement sampling and escalation rules for borderline categories. AI privilege review supports, but does not replace, professional skill, care and judgment. Additionally, a clawback agreement, negotiated with opposing parties before any AI-assisted production begins, is essential.

1. **Disclosure to AI service providers.** General-purpose AI platforms may, under their standard terms, use submitted data to train or fine-tune their models. If a privileged client communication is submitted to such a platform, the confidential content of that communication may become accessible outside the matter. Confidentiality, and therefore privilege, may be irretrievably lost.

As a control, contractual confirmation must be obtained that the platform operates under a zero-retention, no-training policy before any client document is submitted to an AI platform for review or analysis.

1. **Creation of discoverable materials.** AI review platforms generate audit trails, query logs and metadata that may contain significant information about the content of reviewed documents and about the reviewing lawyer's analytical approach. In criminal and regulatory investigations, investigators may specifically seek these AI-generated materials under broad production powers. All statutory compulsory production powers in the United

Kingdom contain a carve-out for privileged material, and the increasing use of AI does nothing to alter that basic position.

As a control, teams should be aware of how such data are created and where they are stored. Teams should treat such data as privileged and ensure they are stored alongside other privileged materials, with restricted access, in a way that makes clear they are artefacts of an activity with a dominant legal purpose.

Drafting legal advice with AI assistance

A memorandum drafted with AI assistance can still attract LAP if it remains a confidential lawyer-client communication created for the dominant purpose of legal advice. Using AI as a drafting aid is conceptually no different from using search tools or word processing technology. However, careful consideration must be given to ensuring confidentiality is maintained and the draft benefits from the lawyer's judgment.

1. **Preserving confidentiality.** Submitting client instructions, privileged emails, interview notes or investigative facts to a general-purpose model can amount to disclosure to a third party. If the platform's terms of service do not impose obligations adequate to maintain confidentiality, privilege may be lost.

To mitigate this risk, teams must only use platforms approved for enterprise-level use, limit inputs to what is necessary, avoid pasting verbatim passages of privileged material into unapproved tools and ensure contractual commitments are in place with the service provider.

1. **Demonstrating legal judgment.** A document consisting entirely of AI-generated text, with no meaningful input from a qualified lawyer, is at greater risk of challenge on two grounds: first, that it was not brought into existence for the purpose of giving legal advice and, second, that the lawyer's involvement was so attenuated that the document cannot properly be characterised as legal advice.

To mitigate this risk, AI-assisted drafts must be critically reviewed, and where appropriate amended, by the lawyer. The final advice document must reflect the lawyer's own professional judgment.

Note-taking and transcription

AI-powered transcription and note-taking tools are now used routinely in corporate environments. When these tools are active during lawyer-client communications, they create an acute privilege risk because they often run by default and transmit data to third-party providers. If they capture privileged lawyer-client communications, privilege may be challenged on the basis that confidentiality was not preserved.

Ideally, at the outset of a matter, the parties should agree that no lawyer-client communications will be recorded. If that is not possible, the parties should agree on the recording and transcription tools that can be used and must ensure that these do not transmit or store data on an external server. Transcripts and recordings should be stored on company systems alongside other privileged materials, rather than stored on the recording application itself.

AI used by third-party service providers

Documents generated by third parties such as forensic accountants and e-disclosure providers may attract LP when created for the dominant purpose of anticipated or ongoing litigation. However, that protection is conditional on confidentiality being maintained. Where a third party processes client data through AI platforms that are not bound by adequate confidentiality obligations, any claim to LP over the third party's work product is weakened.

Before instructing any third-party service provider, practitioners should ensure the following information is incorporated into the engagement letter or a supplementary data processing agreement:

- confirmation of the AI platforms the provider uses;
- the confidentiality terms applicable to those platforms, including data location, retention periods and no-training commitments;
- how prompts, logs, summaries and other derivative artefacts are stored and protected; and
- a commitment that no client-specific information will be processed through unapproved tools and that all AI outputs remain confidential and used only for the retainer.

Agentic AI

Agentic AI systems operate autonomously to complete multi-step tasks with limited human intervention. For example, such systems may automatically review and categorise incoming documents, draft correspondence, extract and summarise data, and update case management records, all without the input of a lawyer. That is substantively and qualitatively different from a lawyer using AI as an effective assistant at each of these steps.

As autonomy increases, so does the risk that outputs are attacked as not being lawyer-client communications (for LAP) or as not being created for the dominant purpose of litigation/legal advice (for LP). The risk is particularly acute where the system's steps and data flows are not documented.

English law has not yet tested how privilege applies to materials generated by autonomous agents. Until it does, firms should treat agentic deployments in investigations as high-risk and ensure instructions given to agentic AI systems are documented in writing, clearly identifying the legal purpose of each task and the name of the supervising lawyer. Firms should also ensure a qualified lawyer reviews and approves all significant outputs generated by agentic systems.

Key recommendations

The actions set out below represent the minimum steps that investigations practitioners and in-house counsel should take to manage privilege risk in the age of AI:

1. **Approved tools register.** Maintain a live list of approved/prohibited tools, configurations and permitted use cases.
2. **AI use protocol.** Define what data may be input, who may use which tools, where outputs/logs live, retention and deletion policies, and escalation routes for borderline cases.
3. **Engagement terms.** Update engagement letters to address AI use by the firm, the client and third parties; include terms regarding confidentiality, retention, no-training commitments and incident notification.
4. **No transcription by default.** Address AI note-taking/transcription explicitly at the outset of matters. Where one party insists on the use of these tools, document the agreement and tool choice before privileged calls.
5. **Require human review.** No AI output should be used, disclosed or relied upon without review and sign-off by a qualified lawyer. This applies across all workflows, including document review, drafting, analysis and agentic tasks.
6. **Clawback arrangements.** Put a clawback mechanism in place before AI-assisted production to mitigate inevitable classification error at scale.
7. **Purpose documentation.** Record the legal purpose and supervising lawyer for significant AI tasks at the time they are initiated. Preserve an audit trail that can be shown if privilege is challenged.
8. **Protect discoverable artefacts.** Treat prompts, logs, summaries and audit trails as sensitive and potentially privileged material, with access controls aligned to other privileged material in the matter.

Authors

Lloyd Firth

PARTNER

lloyd.firth@wilmerhale.com

+44 (0)20 7872 1014

Frederick Saugman

COUNSEL

frederick.saugman@wilmerhale.com

+44 (0)20 7872 1690

1. Three Rivers (No 5) [2003] EWCA Civ 474.

2. Waugh v British Railways Board [1980] AC 521.