

Navigating Legal Privilege Issues When Using AI

By **Charlie Morgan, Maura McIntosh and Charlotte Benton** (December 9, 2025)

The recent explosion in artificial intelligence adoption in the workplace has led to the creation of whole new categories of documents, in the form of prompts and artificial intelligence outputs, as well as more traditional documents that are based on AI outputs.[1]

These documents, like any others, may be susceptible to disclosure in legal proceedings, unless protected by privilege.

Lawyers, both in-house and in private practice, are increasingly using AI to assist in providing legal services, whether through AI-enabled research tools, or to summarize large volumes of documents or streamline the drafting process. And no doubt many individuals in commercial organizations are using AI chatbots to ask questions they might once have asked their legal department or external counsel.

So, it is important to understand when privilege may apply to these interactions with AI systems, and to communications and documents prepared with their assistance. Although these issues have not been addressed by the English courts to date, it seems inevitable that they will be tested in the coming months and years.

In this article we apply familiar principles to assess the likely outcome, while recognizing that predictions cannot be made with complete confidence in such a novel and fast-moving context.

Broader Protection Under Litigation Privilege

Broader protection is likely to be available when litigation is either on the cards or in progress. That is because where litigation is in reasonable prospect, litigation privilege applies to communications or documents created for the dominant purpose of obtaining legal advice, information or evidence relating to the conduct of that litigation.

If this test is met, it is difficult to see why litigation privilege would not apply both to the interactions with an AI system, i.e., the prompt and the AI output, and to any communications or documents created using the AI output.

Importantly, in applying the dominant purpose test, the relevant purpose is that of the person who instigates the creation of the communication or document in question, not necessarily its author.[2] In the AI context, that must be the person inputting the prompt, or creating the communication or document based on the AI output. There is therefore no need to ascribe a purpose to the AI system itself.

The view that litigation privilege can apply to the prompt and the AI output, as opposed to the communication or document ultimately produced, is sometimes questioned on the basis that litigation privilege is said to apply only to communications between lawyer and client,



Charlie Morgan



Maura McIntosh



Charlotte Benton

or between either of them and a third party. This can lead to esoteric arguments about whether an AI system may be considered a third party for these purposes.

But this seems an unnecessary debate. It is of course true that litigation privilege can encompass third-party communications. That is its main advantage over legal advice privilege, which as we discuss applies only to lawyer-client communications.

However, there is no reason to think that litigation privilege is restricted to communications between lawyer and client or either of them and a third party, so that it would not apply to other sorts of communication. As an example, between two members of the client group, or documents that are not communications at all, such as file notes.

Such an approach would seem particularly surprising in circumstances where the Court of Appeal has confirmed that litigation privilege may apply even to nonparties to litigation, such as liability insurers or litigation funders.[3]

Where Litigation Is Not in Prospect

If litigation is not in prospect, the question will be whether legal advice privilege is available. As noted, in contrast to litigation privilege, legal advice privilege applies only to lawyer-client communications or draft communications that are prepared for the dominant purpose of giving or obtaining legal advice.

Other documents or communications will not be covered, unless and to the extent that they evidence the legal advice or the questions on which the lawyer is asked to advise.

On that basis, it seems clear that under current law neither the prompt nor the AI output will be covered by legal advice privilege, save to the extent that they evidence the lawyer's advice or the questions advised upon, and subject to the working papers principle we discuss.[4]

The client who chooses to ask AI for advice, rather than a lawyer, is therefore likely to be out of luck from a legal advice privilege perspective.

However, the use of AI should not affect the application of privilege to lawyer-client communications that have been prepared with the benefit of AI, so long as the dominant purpose test is met and so long as it is a genuine lawyer-client communication. In other words, one that the lawyer or the client has deliberately sent, rather than being automatically generated and dispatched by an autonomous AI agent.

The position may be less certain if the AI output is adopted wholesale to form the lawyer-client communication, but the better view would seem to be that privilege should still attach.

In this scenario, the lawyer or the client will have instigated the production of the AI output by inputting a prompt into the AI system. They will therefore will be responsible for the substance of the communication, even if they did not draft it personally.

They will then have reviewed the AI output and decided to communicate to the client or lawyer in those terms. The version as sent will therefore be a lawyer-client communication for the dominant purpose of seeking or obtaining legal advice.

It is also relevant to note that legal advice privilege applies to the whole continuum of

communications between lawyer and client in a relevant legal context, to enable the lawyer to advise and the client to give instructions as and when needed.[5] It therefore applies to the communication of factual information, including information that is in the public domain.[6]

On that basis, it would seem incongruous if the court were to find that the adoption of AI-generated content in a lawyer-client communication prevented privilege arising.

Lawyers' Working Papers

Where a lawyer is using an AI system, there is a further basis for arguing that both the prompt and the AI output are protected by privilege — namely, that they form part of the lawyer's working papers.[7] There is, however, some uncertainty regarding the extent of this principle.

The High Court decisions in RBS Rights Issue Litigation[8] in 2016 and Serious Fraud Office v. Eurasian Natural Resources Corporation Ltd in 2017[9] suggest that lawyers' working papers are protected by privilege only if and to the extent that their disclosure would betray the tenor of the legal advice.

At appeal in SFO v. ENRC, it was argued that this test is wrong, and that lawyers' working papers will be privileged, so long as they are confidential documents created by a lawyer for the purpose of giving legal advice. However, since the Court of Appeal held that the notes in question were covered by litigation privilege, it did not need to resolve the working papers question.[10]

In any event, in both RBS Rights Issue and SFO v. ENRC, the relevant working papers were lawyers' notes of interviews with individuals who, it was established, were not their clients for the purposes of legal advice privilege.[11] Accordingly, the interviews themselves were not privileged.

Arguably, therefore, the narrow test that these decisions set out for the scope of the working papers principle applies only to lawyers' notes of nonprivileged meetings or discussions. The better view would seem to be that other working papers, including prompts and AI outputs, where these are created by lawyers for the purpose of giving legal advice, will be protected by the principle.

It is worth pointing out that if the position were otherwise, this issue would have a much broader impact than simply the AI context. It would mean that many documents prepared by lawyers for the purpose of giving legal advice could be subject to scrutiny in legal proceedings.

So far as an external lawyer's working papers are concerned, it will often be the case that these are not in the client's control, and so will not be caught by its disclosure obligations. However, the same point cannot be made for the working papers of in-house lawyers.

Meeting Notes Prepared by AI

In general, a transcript or note of a meeting will be privileged only to the extent that the underlying discussion is privileged. So, if litigation is in prospect and the discussion is for the dominant purpose of obtaining legal advice, information or evidence relating to that litigation, it will be covered by litigation privilege.

Otherwise, privilege will attach only to communications from or to any lawyer present at the meeting for the purpose of giving or obtaining their legal advice, or to any aspect of the meeting that reveals the advice or the areas on which the lawyer is advising.

In preparing a note of a meeting that has strayed between privileged and nonprivileged discussions, the key is to ensure that the privileged content is clearly identified — most obviously by making it clear where a lawyer is speaking and where questions or comments are addressed to a lawyer.

In principle, there should be no difference whether the note of the meeting is prepared by a person or by an AI system. If the note is AI-generated, however, there may be less ability to control its contents. It will therefore be important to take care in instructing the AI system, or if necessary to review and amend the draft note to ensure privileged material is properly flagged.

Loss of Privilege

Under English law, confidentiality is a prerequisite for the creation of privilege. Once a privileged document is created, the general principle is "once privileged, always privileged," unless and until privilege is waived or the material ceases to be confidential.[12]

Where a public AI system is used, therefore, there is a risk that privilege will not be available for new material created, or that privilege in any existing material entered into the system will be lost. The obvious message is to use only private AI tools with strict confidentiality protections.

Document Preservation and Disclosure

In general terms, parties are under an obligation to take reasonable steps to preserve potentially relevant documents as soon as litigation is contemplated. At that point, a document hold notice will typically be sent within the organization, instructing relevant teams to stop any routine document destruction procedures and to ensure all those involved in the subject matter of the dispute are aware of the need to preserve documents.

As commercial organizations have embraced new communication technologies, they have had to grapple with fresh challenges around document preservation and disclosure. The use of WhatsApp and other messaging apps has provided an obvious case in point in recent years. The advent of AI presents similar challenges, as it creates another source of documents — in the form of prompts and AI outputs — that may have to be preserved, searched and disclosed if a dispute arises.

The steps that will need to be taken in a given case will depend on the extent to which AI has been used by relevant individuals, and the likelihood that the prompts and outputs may themselves be relevant. This is as opposed to any documents created with the assistance of AI, which should be captured in the usual ways.

Conclusion

As well as opportunities, the widespread adoption of AI brings considerable challenges for businesses and those advising them, including in relation to privilege and document management.

To mitigate these risks, organizations should ensure that robust confidentiality measures

are in place and that staff are educated on the risks of using AI — not least that privilege may not apply where a chatbot is consulted rather than a lawyer. Practices around document preservation and disclosure should also be reviewed to ensure that they keep up with this developing technology.

Charlie Morgan is a partner, Maura McIntosh is a knowledge counsel and Charlotte Benton is a senior associate at Herbert Smith Freehills Kramer LLP.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] In this article we are using the term "artificial intelligence" or "AI" to refer to generative AI or GenAI, rather than more traditional or rule-based AI systems or machine learning.

[2] See for example *Victorygame Ltd v. Ahuja Investments Ltd.* [2021] EWCA Civ 993.

[3] *Al Sadeq v. Dechert LLP* [2024] EWCA Civ 28.

[4] It is conceivable that privilege could be extended to cover legal advice provided by AI systems that have been specifically designed for that purpose, but that would require a development of the current law that requires a lawyer/client communication.

[5] *Balabel v. Air India* [1988] Ch 317.

[6] *Property Alliance Group Ltd v. The Royal Bank of Scotland PLC* [2015] EWHC 3187 (Ch).

[7] This was one of the categories of documents for which a claim to privilege was upheld in *Balabel v. Air India* [1988] Ch 317.

[8] *The RBS Rights Issue Litigation* [2016] EWHC 3161 (Ch).

[9] *SFO v. Eurasian Natural Resources Corporation Ltd.* [2017] EWHC 1017 (QB).

[10] *SFO v. Eurasian Natural Resources Corporation Ltd.* [2018] EWCA Civ 2006.

[11] And there was no litigation in contemplation, so litigation privilege was not available.

[12] See for example *Addlesee v. Dentons Europe LLP* [2019] EWCA Civ 1600.