INSIGHTS

Media & Entertainment

The Impact of Using Generative AI on Maintaining Privilege

Early judicial decisions highlight risks of inputting confidential legal materials into a generative AI platform

By James Rosenfeld, Elyse Sparks, and Dalia Wrocherinsky

02.27.26

Technology companies have raced to create AI-driven legal tools that bring both tremendous efficiency and great risk, resulting in a slew of sanctions rulings assessed against litigants relying on AI-fabricated legal citations.[1] The impact of AI on privilege has not drawn as much attention. Can a party's exchanges with a generative AI platform during a litigation be protected under the attorney-client privilege or work product doctrine? Can inputting them constitute a waiver of these protections? Courts have begun to grapple with these issues, reaching interesting and inconsistent results. Two early decisions highlight both the potential risks and the uncertain state of the law.

U.S. v. Heppner

On February 17, 2026, Judge Jed Rakoff of the U.S. District Court for the Southern District of New York issued an opinion addressing whether communications between a criminal defendant and a publicly available generative AI platform are protected by the attorney-client privilege or the work product doctrine. In *United States v. Heppner*, No. 25 Cr. 503 (S.D.N.Y. Feb. 17, 2026), the court held that they are not rejecting privilege claims over documents the defendant created through interactions with Anthropic's "Claude" platform while under criminal investigation.

The defendant, an executive charged with securities fraud and related offenses, used Claude in anticipation of being indicted after receiving a grand jury subpoena. He generated written "reports" outlining potential defenses and legal arguments, later sharing those materials with counsel. In exchanges with the government, defendant's counsel later asserted privilege over these documents, arguing that defendant had (1) inputted information learned from counsel, (2) created the documents for the purpose of obtaining legal advice from counsel, and (3) subsequently shared the AI's output with counsel. Defendant conceded, however, that counsel had not directed him to run the Claude searches.

The government sought a ruling that the AI-generated materials were neither attorney-client communications nor protected work product. The court agreed, granting the government's motion.

On the issue of attorney-client privilege, the court focused on the three core requirements of the privilege: (1) a communication between client and attorney; (2) which is intended to be, and is, confidential; and (3) for the purpose of obtaining or providing legal advice. The court held that the AI exchanges failed on "two, if not all three" of the requirements:

- First, the communications were between the defendant and an AI platform, not between the defendant and an attorney. The court rejected the argument that using AI is like using internet-based word processing software because the use of such applications is not intrinsically privileged and, as software, does not involve a "trusting human relationship" like the attorney-client relationship.
- Second, the communications were not confidential. The court emphasized that defendant voluntarily disclosed inputs to a third-party platform whose privacy policy permits data retention and disclosure of users' inputs and Claude's outputs. Because Anthropic reserves the right to disclose such data to third parties, including government authorities, and provides notice to its users of such, the court held that the defendant could have had "no reasonable expectation of confidentiality in his communications."
- Third, the defendant did not consult the AI for the purpose of obtaining legal advice.[2] Despite defendant's claim that he used Claude for the express purpose of talking to counsel, he did not do so at the direction of counsel—which the court suggests may have made the difference. Moreover, the court noted that later sharing the resulting materials with counsel did not "alchemically" transform them into privileged communications.

The court likewise rejected work product protection. The materials were not prepared "by or at the behest of counsel," and they did not disclose nor affect counsel's mental impressions or strategy. Because the defendant acted on his own initiative in using the AI tool, the court held that the documents fell squarely outside of the doctrine's core purposes.

Warner v. Gilbarco

Just one week prior to Heppner, however, a less widely reported ruling in *Warner v. Gilbarco* (No. 2:24-cv-12333) from the Eastern District of Michigan, demonstrated that a different set of facts in front of a different judge may lead to an entirely different outcome. In *Warner*, the defendant sought to compel the plaintiff to produce all materials related to her use of ChatGPT in connection with the lawsuit, similarly arguing that information fed into a generative AI tool was not protected by the work product privilege, or amounted to a waiver of that privilege.

The court rejected that theory on the ground that materials prepared by a pro se litigant in anticipation of litigation were protected by the work product doctrine, and waiver requires disclosure *to an adversary* (or at least in a manner likely to end up in an adversary's hands). The court held that "ChatGPT (and other generative AI programs) are *tools*, *not persons*, even if they may have administrators somewhere in the background," and thus sharing mental impressions with a tool is fundamentally different from handing them to opposing counsel. Accordingly, for this court, concluding that the use of generative AI waives work-product protection would "nullify work-product protection in nearly every modern drafting environment."

Takeaways From These Cases

1. **Train employees on privilege in the AI context:** Companies should provide targeted training not only to executives and in-house counsel, but to any employee who may be involved in litigation, investigations, regulatory inquiries, or internal reviews. Employees should understand when privilege attaches, how it can be waived, and why using public AI tools to analyze legal exposure or summarize legal advice may create discoverable materials. In a post *Heppner* world, privilege preservation depends on clear guidance and consistent training across the organization.
2. **Rely on the direction of counsel:** As seen in *Heppner*, AI-generated materials created on a client's own initiative, even if later shared with counsel, are at a high risk of being unprotected. To maximize the chance of privilege protection, AI use in connection with litigation or legal matters should be directed by counsel, integrated into the attorney's workflow, and documented as such.
3. **Negotiate enterprise terms instead of standard privacy policy:** The *Heppner* court placed significant weight on the fact that Anthropic's terms permitted data retention and disclosure, which defeated any reasonable expectation of confidentiality. Corporate clients should closely review their agreements with AI providers and, where possible, negotiate for enterprise terms that include robust data confidentiality protections, prohibit training on user inputs, and limit storage, use, and third-party disclosure.
4. **The law is unsettled and fact specific:** *Heppner* and *Warner* reached different conclusions on overlapping questions, one week apart. Until courts reach consensus, companies should bear in mind that exchanges with generative AI may be unprotected, at least unless counsel is actively involved. Internal AI use policies should reflect this uncertainty and be revisited as the case law develops.

+++

Jim Rosenfeld is a partner in the media group in the New York office of DWT. Elyse Sparks and Dalia Wrocherinsky are associates in the technology transactions group in the Seattle office of DWT.

---

[1] Jennifer Kay, *Fake AI Citations Produce Fines for California, Alabama Lawyers*, Bloomberg Law (Oct. 13, 2025).

[2] Since defendant acted of his own volition, the court indicated that the pertinent question for this factor was whether he was seeking legal advice *from Claude*. Ironically, it looked to Claude itself for the answer: when prompted for legal advice, the court noted, Claude responded that it was not a lawyer and could not provide legal advice, directing the user to consult with a qualified attorney.

Related Articles

02.17.26

INSIGHTS

Intellectual Property

The Public Domain Shakedown: Paying for Rights No One Owns

JANUARY 2026

PUBLICATIONS

Artificial Intelligence

Agentic AI Payments: Navigating Consumer Protection, Innovation, and Regulatory Frameworks

01.23.26

INSIGHTS

Media & Entertainment

Wave of Federal "Online Safety" Legislation Hits Congress