Insights: Alert

Protective Orders in the Age of Generative AI: Best Practices for Safeguarding Confidential Information

May 1, 2026

Written by **Joel D. Bush**

Federal courts are rapidly developing protective order language to address a new and significant risk: that confidential discovery materials will be exposed to generative AI platforms that retain user inputs, use them for training, or share them with third parties. Between 2025 and early 2026, at least five federal district courts have issued or analyzed specific protective order provisions restricting the use of generative AI platforms with protected materials, and a federal appellate court has underscored attorneys' ethical obligations when using these tools. These decisions establish that proactive AI-specific provisions are no longer optional — they must be standard practice.

The Legal Framework

Federal Rule of Civil Procedure 26(c)(1)(G) authorizes courts to issue protective orders for good cause requiring that "a trade secret or other confidential research, development, or commercial information not be revealed or be revealed only in a specified way." This provision serves as the central statutory basis for AI-specific restrictions in protective orders, and courts have applied it with the "highly flexible" good-cause standard described in *Rohrbough v. Harris*, 549 F.3d 1313, 1321 (10th Cir. 2008).

The foundational concern is straightforward: publicly available generative AI platforms typically retain user inputs, use submitted data for model training, and reserve the right to disclose data to third parties, including governmental regulatory authorities. As the Southern District of New York held in *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026), a user of a publicly available AI platform "could have had no 'reasonable expectation of confidentiality in his communications' with Claude," given Anthropic's express privacy policy permitting data collection, training use, and third-party disclosure. According to the court in *Heppner,* submitting privileged information to such a platform constitutes a third-party disclosure that waives attorney-client privilege entirely, "just as if he had shared it with any other third party."

Beyond privilege waiver, courts have recognized that the practical impossibility of retrieving data once submitted to open AI tools creates distinct risks. In *Jeffries v. Harcros Chemicals Inc.*, 2026 WL 820218 (D. Kan. Mar. 25, 2026), the court explained that open AI tools "risk disclosure, loss of control, and uncertainty concerning the security, storage, and other data-handling of that information," and that it is "practically impossible" to claw back data from an open AI tool because "the information was used to train the AI Tool." The court also flagged potential violations of U.S. data privacy laws and the European General Data Protection Regulation (GDPR).

AI as “Tool” vs. AI as “Third Party”: The Privilege/Work Product Distinction

Courts have drawn an important distinction between how AI platform use is treated under the attorney-client privilege versus the work product doctrine.

For attorney-client privilege purposes, *Heppner* treats AI platform inputs as disclosures to a third party, destroying confidentiality and defeating privilege. For work product purposes, however, the analysis is different. The Eastern District of Michigan in *Warner v. Gilbarco, Inc.*, 2026 WL 373043 (E.D. Mich. Feb. 10, 2026) held that public generative AI programs are “tools, not persons, even if they may have administrators somewhere in the background,” and that work product waiver “has to be a waiver to an adversary or in a way likely to get in an adversary's hand.” The court in *Warner* found no work product waiver from a pro se plaintiff's use of generative AI tools in preparing litigation materials.

Morgan v. V2X, Inc., 2026 WL 864223 (D. Colo. Mar. 30, 2026) similarly held that “AI interactions do not automatically compromise work product protections” because it is “highly unlikely the information will fall into the hands of an adversary absent some legal process to compel it.” The *Morgan* court distinguished *Heppner*, noting that in the pro se context “[a] pro se litigant is simultaneously the party and the advocate,” eliminating the gap between party and attorney that was dispositive in *Heppner*.

This distinction is consequential: while work product protection may survive AI platform use, attorney-client privilege clearly does not — and neither does the confidentiality of materials designated as protected under a protective order if they are submitted to platforms that retain data rights.

Four Emerging Approaches to AI Protective Order Language

Federal courts have developed four principal models for addressing generative AI disclosure risks in protective orders.

Model 1: The Blanket Prohibition. In *Fiorito v. Metropolitan Council*, 2025 WL 1806612 (D. Minn. 2025), the court entered a protective order providing:

> No party may upload or otherwise disclose Confidential Data to any generative artificial intelligence platform, such as (but not limited to) ChatGPT.

This language is brief, clear, and easy to apply, but it does not distinguish between open consumer platforms (which present serious confidentiality risks) and enterprise or closed AI tools that may offer strong contractual data protections.

Model 2: The Data-Rights Approach. In both *United States v. Jackson*, 2025 WL 1666249 (W.D. Wash. 2025) and *United States v. Navarro Hernandez*, 2025 WL 3290753 (W.D. Wash. 2025), the court entered protective orders providing:

> The Protected Material shall not be loaded into an artificial intelligence (AI) tool or model, generative or not, where rights to the data are retained by any party that has not explicitly agreed to be covered by the protective order.

This formulation applies to AI tools of “any” kind — generative or otherwise — and focuses on whether the AI provider has agreed to be bound by the protective order's requirements, permitting use of AI tools that commit

contractually to data protections aligned with the order.

Model 3: The Contractual Safeguards Approach. The most detailed framework was crafted by the District of Colorado in *Morgan v. V2X*. The court rejected both parties' proposals and entered its own language:

> No party or authorized recipient may input, upload, or submit CONFIDENTIAL Information into any modern artificial intelligence platform, including any generative, analytical, or large language model-based tool ('AI'), unless the AI provider is contractually prohibited from: (1) storing or using inputs to train or improve its model; and (2) disclosing inputs to any third party except where such disclosure is essential to facilitating delivery of the service. Where disclosure to a third party is essential to service delivery, any such third party shall be bound by obligations no less protective than those required by this Order. In addition, the AI provider must contractually afford the party or authorized recipient the ability to remove or delete all CONFIDENTIAL information upon request. A party intending to use AI that it contends meets these requirements must retain written documentation of these contractual protections.

The court observed that the practical effect is to bar use of “most, if not all, mainstream low-to-no-cost AI to process Confidential Information,” including “standard ChatGPT, Claude, Gemini, or similar platforms.”

The *Morgan* court rejected the plaintiff's proposed language permitting uploads to AI operating in a “secure, closed-circuit environment,” finding that it addressed cybersecurity concerns about unauthorized access “rather than the distinct risks associated with today's widely available AI platforms.” It also rejected the defendant's proposed language as “over-engineered” and “crafted to fit the precise bounds of Defendant's contractual engagement with AI providers,” resulting in provisions that were “vague” from the opposing side's perspective. The court further cautioned parties against over-designating information as “Confidential” in light of this restriction and ordered the plaintiff to disclose the name of any AI platform previously used with confidential materials.

Model 4: The Notice-and-Secure-Environment Approach. *Jeffries v. Harcros Chemicals Inc.* presents the most comprehensive AI framework. The original protective order required that a party intending to use an AI tool must:

> (1) provide other parties with notice and an opportunity to object;
>
> (2) ensure the AI tool is used in a secure environment and that confidential information is not used to train or improve any AI tool except one used exclusively in the action;
>
> (3) ensure that any confidential information used to train an AI tool is destroyed at the conclusion of litigation and not made accessible to unauthorized persons; and
>
> (4) ensure all confidential information is deleted at the conclusion of the action.

The court subsequently granted a motion to expand these AI restrictions to all discovery materials — including non-confidential documents — prohibiting their input into “open loop” or public generative AI tools entirely.

The court found good cause based on the practical impossibility of clawing back data from open AI tools, security risks of a centralized repository at unprecedented scale, potential GDPR violations, and the critical infrastructure designation of some defendants' operations. Additionally, the court rejected First Amendment objections, noting the order was limited to the pretrial discovery context and did not prevent public dissemination of non-confidential information from other sources.

Attorney Ethical Obligations

The Sixth Circuit in *United States v. Farris*, 171 F.4th 920 (6th Cir. 2026) underscored that "attorneys who choose to use artificial-intelligence tools must do so in a manner consistent with their ethical obligations," specifically identifying "safeguarding confidential client information and preserving attorney-client privilege" as a relevant step. The court referenced the ABA Task Force on Law & AI's Year 2 Report (2025) as compiling applicable ethics rules and guidance across jurisdictions. These ethical obligations exist independently of any protective order and reinforce the need for practitioners to address AI platform use expressly in proposed protective orders.

In *Farris*, the court sanctioned an attorney who used AI to draft appellate briefs without properly verifying the cited legal authorities, resulting in false quotations and misleading legal arguments. The court ordered that counsel not be compensated under the Criminal Justice Act, forwarded the opinion to disciplinary authorities, and appointed replacement counsel. The case is a sobering reminder that competence and candor obligations "apply no matter the tools attorneys use."

What You Should Be Doing Now

Given the rapid development of this area, practitioners should take the following steps:

Proactively propose AI-specific provisions. Courts have demonstrated a willingness to impose AI restrictions even when the parties have not initially included them. The most protective approach — and the one least likely to generate disputes — is the contractual safeguards model from *Morgan v. V2X* [Model 3 above], which provides a clear standard applicable to any AI tool regardless of platform.

Identify compliant AI tools before litigation begins. Parties who intend to use AI for document review, legal research, or other litigation tasks involving protected materials should identify enterprise-tier AI products that meet contractual safeguard requirements — including contractual prohibitions on training and third-party disclosure, deletion rights, and written documentation — before the dispute arises.

Never submit privileged information to publicly available AI platforms. Even absent a protective order, submitting privileged information to publicly available AI platforms risks waiver of attorney-client privilege under *Heppner*.

Assess whether existing protective orders already cover AI disclosures. Where a protective order is already in place with broad disclosure prohibitions, submitting protected materials to AI platforms that retain data rights may independently violate the order, as the *Morgan* court observed.

Review and update internal policies. Educate clients and teams on privilege risks and best practices regarding AI use. Implement internal guidelines restricting the use of generative AI for privileged or confidential work unless protected by adequate safeguards.

Consider the expanded *Jeffries* approach for sensitive cases. In cases involving critical infrastructure, GDPR-subject entities, or other heightened data sensitivity, consider proposing restrictions covering all discovery materials — not just those designated as confidential — given the *Jeffries* court's willingness to expand AI restrictions to the full scope of produced discovery.

Retain documentation of AI contractual protections. Under the *Morgan* framework, a party intending to use AI must retain written documentation of contractual protections. Practitioners should maintain these records as a matter of course and be prepared to identify any AI platforms used with confidential materials to the court.

We will continue to monitor the evolving case law and guidance from courts regarding protective orders and AI use. Please contact us if you have questions about how these decisions may affect your organization.

Related People

Joel D. Bush

t 404.815.6074

jbush@ktslaw.com