THE HEPPNER PROTECTIVE SHIELD (HPS)

A Framework for Preserving Attorney-Client Privilege in the Era of Agentic AI

By Roland G. Ottley, Esq., PA-C *Principal Attorney | The Ottley Law Firm, P.C. Federal Civil Litigator | Licensed Physician Assistant | Author, Agentic Fidelity (AgFi) Framework*
1063 Winthrop Street, Brooklyn, New York 11212 | rottley@tolfpc.com | (718) 221-2162
March 15, 2026

EXECUTIVE SUMMARY

On February 10, 2026, Judge Jed S. Rakoff of the United States District Court for the Southern District of New York issued a bench ruling—memorialized in a February 17, 2026 written opinion—holding that 31 documents a defendant generated using the consumer version of Anthropic's Claude were not protected by attorney-client privilege or the work product doctrine. *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026). [1]

This was not a revolution in privilege law. It was a precise application of settled doctrine to a fact pattern shaped by one decisive variable: the defendant acted alone, without counsel's direction, on a platform whose consumer privacy policy expressly permits disclosure to governmental regulatory authorities.

This article provides the complete legal analysis and factual context of *Heppner*; corrections to misreadings circulating in bar publications; a medical-legal analogy illuminating HIPAA risks attorneys are missing; a platform-by-platform confidentiality analysis with verified pricing data for seven major AI platforms; integration of the TOLFPC Agentic Fidelity (AgFi) framework as a privilege predicate; comparative analysis of general versus specialized legal AI platforms; a comprehensive survey of the

2026 regulatory landscape; a full Practical Roadmap for the safe adoption of agentic AI; and the complete TOLFPC Three-Part Anti-Hallucination Protocol—the minimum standard of competence for AI-assisted legal work.

Crucially, this manuscript formalizes the **Heppner Protective Shield (HPS) Doctrine**, a testable, operational framework designed to preserve attorney-client privilege and work product protections when utilizing agentic AI systems. By synthesizing the *Kovel* doctrine, the *Bieter* functional equivalent doctrine, the *Ackert* limitation, and modern technological realities, the HPS Doctrine provides courts and practitioners with a clear standard for evaluating AI-assisted legal work.

Top-Line Conclusion: If an attorney uses an AI platform that provides genuine, contractually enforceable privacy, confidentiality, and non-retention protections, then —under the *Kovel* doctrine, the functional equivalent exception, and the prevailing jurisprudence in the Sixth, Seventh, Eighth, and Ninth Circuits—the attorney's prompts are likely protected by attorney-client privilege and the work-product doctrine. The key determinant is whether the AI provider is treated as a functional agent of the attorney, not a third-party recipient. The 2026 Michigan case (*Warner v. Gilbarco*) strongly supports protection when confidentiality safeguards are in place, while the competing SDNY *Heppner* decision shows the opposite outcome when such safeguards are absent.

The Central Thesis: The legal profession is currently operating under a dangerous misconception that the attorney-client privilege automatically extends to generative AI platforms simply because they are used for legal work. This article argues that under the *Heppner* precedent, the privilege does not survive contact with an AI platform unless the attorney proactively constructs a specific, verifiable Heppner Protective Shield (HPS). The HPS Doctrine posits that an AI platform is legally indistinguishable from a third-party human consultant under the *Kovel* and *Bieter* doctrines. Therefore, privilege is preserved *only* when the AI is deployed under strict attorney direction, governed by an enterprise-grade Zero Data Retention (ZDR) architecture, and utilized as a necessary conduit for legal advice rather than an independent source of expertise. Without this architecture, the use of AI constitutes a voluntary disclosure to a third party, resulting in an absolute waiver of privilege. The HPS Doctrine is designed to be operational rather than merely aspirational. It provides a testable, litigation-ready standard that courts can apply today, using existing doctrinal tools, to resolve the most consequential privilege question of the digital age.

I. THE FACTS THAT ACTUALLY DROVE THE RULING

Before analyzing what *Heppner* means, attorneys must understand what *Heppner* actually involved—because the most dangerous misreading circulating in client alerts is the suggestion that using AI for legal work is inherently incompatible with privilege. It is not. The ruling turned on a specific, and remarkably avoidable, set of facts.

Bradley Heppner was the former Chief Executive Officer of Beneficient, a Dallas-based alternative asset management firm, and GWG Holdings, a publicly traded specialty finance company. He was charged in the Southern District of New York with securities fraud, wire fraud, and falsification of corporate records arising from an alleged scheme to defraud investors of approximately $150 million. His defense was handled by Quinn Emanuel Urquhart & Sullivan, LLP. Trial was scheduled for April 6, 2026.

After receiving a grand jury subpoena and retaining counsel, Heppner—acting entirely on his own initiative, without any direction from his attorneys—used the consumer (non-enterprise, non-API) version of Anthropic's Claude to generate 31 documents. Those documents analyzed his potential defenses, the government's likely charges, and the facts he believed relevant to his case. Critically, he inputted into Claude information he had received in privileged conversations with his attorneys at Quinn Emanuel.

He then transmitted the AI-generated documents to his lawyers. When the FBI executed a search warrant at his residence, agents seized devices containing these documents. Defense counsel asserted both attorney-client privilege and work product doctrine protection. The government moved to compel production. Judge Rakoff granted the government's motion from the bench on February 10, 2026, and issued his written opinion one week later. [1]

Key Holding — Three Independent and Each Independently Fatal Failures

The Court's decision rested on three independent failures, each of which was independently sufficient to destroy both the privilege and work product claims:

First, Judge Rakoff stated: "Heppner does not, and indeed could not, maintain that Claude is an attorney." [1] The opinion clarified that in the absence of an attorney-client relationship, discussions about legal issues between two non-attorneys are not

privileged. The Court emphasized that privileges require a "trusting human relationship" with a licensed professional, which is absent with AI platforms. [1]

Second, the communications lacked confidentiality. Anthropic's Privacy Policy, to which users consent, explicitly states that user inputs and Claude's outputs can be retained, used for model training, and disclosed to third parties, including "governmental regulatory authorities." [1] Judge Rakoff concluded that Heppner could have no "reasonable expectation of confidentiality in his communications" with Claude given these terms. The Court reiterated that non-privileged communications do not become privileged merely by being shared with counsel.

Third, the communications were not made for the purpose of obtaining legal advice. Claude itself disclaims providing legal advice, and Heppner's use of the platform was not at the direction or suggestion of his counsel. [1]

Regarding the work product doctrine, the Court also rejected its application. The AI-generated documents were not prepared by or at the direction of counsel, nor did they reflect counsel's mental processes or strategy. Heppner's counsel conceded that the documents were prepared "on his own volition" and did not reflect counsel's strategy at the time of creation. [1]

Critical Clarification: The consumer policy's governmental disclosure provision—not merely the training data opt-in—was the dispositive confidentiality failure. As of September 2025, Anthropic gave consumer users the ability to opt out of model training, but the governmental disclosure provision remained in the consumer terms. Enterprise agreements eliminate this provision entirely.

I.B. THE LITERATURE GAP: MOVING BEYOND THEORETICAL ETHICS TO OPERATIONAL DOCTRINE

The existing literature on artificial intelligence in legal practice has largely bifurcated into two distinct, and ultimately insufficient, camps. The first camp focuses on the ethical implications of AI hallucination, relying heavily on the American Bar Association's Model Rule 1.1 (Competence) and Rule 1.6 (Confidentiality). Scholars in this camp, such as those analyzing the *Mata v. Avianca* sanctions, correctly identify the risks of unverified AI outputs but fail to provide a structural legal defense for when those outputs are inevitably subpoenaed. [14] The second camp focuses on the

technical mechanics of prompt engineering and the statistical probabilities of large language model (LLM) accuracy, as seen in the empirical studies by Magesh et al. (2025) and Dahl et al. (2024). [11] [12] While these studies are vital for understanding the limits of the technology, they treat the AI as a black box rather than a legal entity subject to the rules of evidence and discovery.

What is entirely missing from the current literature is a unified doctrinal framework that bridges the gap between technical architecture and evidentiary privilege. Prior scholarship has treated AI as a novel, unprecedented phenomenon requiring entirely new laws. This article argues the opposite: the tools to protect AI-assisted legal work already exist within the established jurisprudence of third-party agency, specifically the *Kovel* and *Bieter* doctrines. By mapping the technical realities of enterprise AI—such as Zero Data Retention and API architecture—directly onto the legal requirements of the functional equivalent exception, this article fills the critical gap in the literature, providing the first operational, litigation-ready framework for defending AI use in federal court.

Furthermore, the existing literature has failed to grapple with the evidentiary dimension of AI outputs. Scholars have debated whether AI-generated legal research is competent under Rule 1.1, but no prior article has systematically addressed whether AI-generated analysis can survive a *Daubert* challenge under Federal Rule of Evidence 702. This article addresses that gap directly in Section XIV, providing the first doctrinal analysis of the *Daubert* vulnerability of black-box AI and the human-in-the-loop evidentiary bridge required to admit AI-generated analysis in federal court.

II. WHAT THE RULING DOES NOT HOLD—CORRECTING THE MISREAD

Numerous law firm client alerts have overstated this ruling's reach. The written opinion is unambiguous: this was a technology-neutral application of settled privilege doctrine to a specific set of facts. Judge Rakoff expressly declined to hold that AI use is categorically incompatible with privilege or work product protection. [1]

A. The Kovel Opening Is Wide

The most underreported and arguably most significant aspect of Heppner is what Judge Rakoff did said about what might have preserved privilege. The court stated that had counsel directed Heppner to use Claude, it "might arguably" have functioned as a lawyer's agent within the protection of the attorney-client privilege—analogizing to the *Kovel* doctrine. [1]

In *United States v. Kovel*, 296 F.2d 918 (2d Cir. 1961), Judge Henry Friendly held that an accountant employed by a law firm to assist counsel in rendering legal advice could be within the privilege, provided the purpose was to assist the attorney in giving legal advice, not to provide independent tax advice. [2] The Second Circuit's *Kovel* framework has since been extended to interpreters, investigators, experts, and other third-party professionals who serve as agents of counsel.

The implication for AI is direct and actionable: enterprise AI tools deployed at counsel's direction, under a contractual framework prohibiting data use for training and disclosure to third parties, in service of rendering legal advice, may fall squarely within *Kovel* protection. *Heppner* did not close this door. It left it wide open.

However, as the *Ackert* limitation makes clear, "wide open" is not the same as "certain." In *United States v. Ackert*, 169 F.3d 136 (2d Cir. 1999), the Second Circuit held that privilege does not extend to a third party who provides independent expertise rather than merely translating client communications for counsel. [3] Applied to AI, this creates the **Ackert Fracture Line**: when an AI organizes and translates client documents for counsel's review, it functions as a *Kovel* agent. When it generates novel legal theories from its training data, it provides independent expertise—and *Ackert* may exclude that output from privilege. Courts have not yet resolved this distinction in the AI context, but it is the most significant unresolved doctrinal question in AI privilege law.

B. The Bieter Doctrine — The Functional Equivalent Exception

Beyond *Kovel*, the *Bieter* doctrine (from *In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994)) stands for a foundational principle that is highly relevant to the AI context: the "functional equivalent" exception to third-party waiver.

The Bieter Rule: Disclosure of privileged information to a third party does not waive privilege if the third party is acting as the "functional equivalent" of an employee or

agent necessary to facilitate legal representation.

Key Components of the Functional Equivalent Exception:

- The third party must be necessary or highly useful to the legal task.
- The disclosure must be reasonably necessary to obtain legal advice.
- The third party must be subject to strict confidentiality obligations (contractual, regulatory, or functional).
- The attorney must have a reasonable expectation of privacy.

Why Bieter Matters for AI: If an AI platform is under a strict confidentiality agreement, does not retain prompts, does not use data for training, and functions as a tool assisting the attorney, then under the functional equivalent doctrine, the AI provider is treated like a translator, a paralegal, an e-discovery vendor, or a cloud-storage provider. In that case, privilege is not waived.

C. The Confidentiality Finding Was Policy-Specific

Judge Rakoff's confidentiality analysis was grounded in the specific terms of Anthropic's consumer privacy policy. Enterprise and API tiers of Claude provide zero data retention (ZDR) agreements, prohibit use of client inputs for model training, and do not contain the governmental disclosure provision that destroyed confidentiality in *Heppner.* [1] An attorney using Claude Enterprise under a properly executed Data Processing Agreement operates in a factual universe that *Heppner*'s holding does not govern.

This is not a distinction between AI platforms. It is a distinction between contractual tiers of the same platform.

D. The Parallel Case: Warner v. Gilbarco (E.D. Mich. Feb. 10, 2026)

On the same date as Judge Rakoff's bench ruling, a federal court in the Eastern District of Michigan reached a highly protective result regarding AI use in litigation. In *Warner v. Gilbarco, Inc.*, No. 2:24-cv-12333, 2026 WL 373043 (E.D. Mich. Feb. 10, 2026), the Warner court held plaintiff's ChatGPT interactions for legal research and document organization constituted protected work product. [4]

This is arguably one of the most significant modern cases on AI and work product defendants moved to compel production of the plaintiff's AI interactions, arguing that

inputting litigation materials into a public AI platform waived protection. The court rejected this argument, holding that the AI-generated materials were protected under Federal Rule of Civil Procedure 26(b)(3)(A), which shields materials "prepared in anticipation of litigation or for trial by another party or its representative." Because a *pro se* litigant is the party, her materials qualified under the plain text of the rule.

Crucially, the court rejected the discovery request as a "fishing expedition" seeking the plaintiff's "internal analysis and mental impressions." The court noted that "the work product waiver has to be a waiver to an adversary or in a way likely to get in an adversary's hand." Most significantly for the legal profession, the court characterized the defendants' theory as one that, "if accepted, would nullify work-product protection in nearly every modern drafting environment, a result no court has endorsed." [4]

The divergence between *Heppner* and *Warner v. Gilbarco* is analytically significant. In *Heppner*, the court found no protection because the defendant acted outside the scope of legal representation on a platform that permitted governmental disclosure. In *Warner*, the court recognized that modern drafting environments—even those utilizing AI—do not automatically waive work product protection simply because a third-party tool is used to process the information.

F. Tremblay v. OpenAI: Attorney Prompts as Opinion Work Product

In *Tremblay v. OpenAI, Inc.*, 2024 WL 3748003 (N.D. Cal. Aug. 8, 2024), the district court granted relief from a magistrate's discovery order, finding that unused ChatGPT prompts "were queries crafted by counsel and contain counsel's mental impressions and opinions about how to interrogate ChatGPT, in an effort to vindicate Plaintiffs' copyrights against the alleged infringements." [5] The court held these unused prompts constituted protected opinion work product—materials that reveal the attorney's mental impressions, conclusions, and legal theories. However, prompts that were actually used and disclosed in the complaint had to be produced. [5]

This holding is significant because it identifies a category of AI-assisted materials that receives near-absolute work product protection: not the AI's output, but the attorney's prompt design. An attorney who structures a precise, theory-specific prompt to direct an AI through a complex legal analysis has embedded their professional judgment in that prompt structure—and courts applying *Tremblay* would protect the prompt from discovery as opinion work product regardless of what the AI produced in response.

E. The Krafton Distinction: Discoverability vs. Privilege

The critical distinction between privilege preservation and mere discoverability risk is starkly illustrated by the Delaware Chancery Court's recent decision in *Fortis Advisors, LLC v. Krafton, Inc.*, C.A. No. 2025-0805-LWW (Del. Ch. Mar. 16, 2026). [71] While *Heppner* serves as the definitive warning regarding privilege waiver, *Krafton* demonstrates the devastating evidentiary consequences of consumer AI use in corporate strategy—a scenario entirely outside the scope of the HPS Doctrine but highly relevant to the broader necessity of enterprise AI safeguards.

In *Krafton*, the parent company's CEO, Changhan Kim, utilized ChatGPT to formulate a "takeover strategy" to oust the founders of subsidiary Unknown Worlds Entertainment and avoid a $250 million earnout payment related to the video game *Subnautica 2*. [72] When internal executives warned that firing the founders without cause might trigger a lawsuit, Kim "turned to ChatGPT for help." [73] The AI chatbot prepared a "Response Strategy to a 'No-Deal' Scenario," including a "pressure and leverage package" and an "implementation roadmap by scenario," which Krafton subsequently executed. [74] Vice Chancellor Will found that Krafton's justifications for the terminations were "pretextual" and that Krafton had breached the Equity Purchase Agreement, ordering reinstatement of the studio's CEO and extending the earnout period.

Crucially, *Krafton* is fundamentally distinguishable from *Heppner* and the HPS framework on three core doctrinal and operational grounds:

1. Core Legal Issue (Substantive Evidence vs. Privilege)

In *Heppner*, the dispute centered squarely on whether AI-generated documents incorporating privileged communications qualified for attorney-client privilege or work-product protection. The entire HPS Doctrine—including the *Kovel*/*Bieter* analysis, the Ackert Fracture Line, and the Confidentiality Architecture gate—is designed to solve this exact privilege-waiver problem. In *Krafton*, no privilege or work-product claim was ever asserted over the ChatGPT conversations. The logs were treated as ordinary business records and admitted as substantive evidence to prove premeditation and bad faith breach of the Equity Purchase Agreement. The court never analyzed *Kovel*, *Bieter*, or privilege doctrine.

2. Attorney Direction Element

The HPS Doctrine requires that AI use be strictly directed by counsel under enterprise-tier safeguards. In *Heppner*, the defendant acted "entirely on his own initiative, without any direction from his attorneys," triggering a fatal failure under the HPS test's Attorney Direction element. In *Krafton*, the user was the company's CEO acting unilaterally for internal corporate strategy. No outside counsel directed, supervised, or even knew about the ChatGPT sessions at the time they occurred. The HPS "Attorney Direction" and "Privilege Documentation (The Tremblay Element)" pillars simply do not apply to the

Krafton fact pattern.

3. Purpose of the AI Use

The HPS framework addresses the use of AI to analyze legal defenses and process privileged attorney communications for the purpose of obtaining legal advice. In *Krafton*, the AI was used for pure business and corporate planning—contriving a takeover strategy, planning executive terminations, and locking Steam platform access. The manuscript's privilege-focused analysis, including the *Kovel* agency doctrine, the work-product doctrine, and the anti-hallucination protocol for legal drafting, has no doctrinal application to such business use.

The Practical Takeaway: Complementary Warnings

Both *Heppner* and *Krafton* involved consumer-tier AI platforms whose terms permitted data retention and disclosure, and both involved logs that were recovered from the provider. However, while this destroyed privilege in *Heppner*, it simply made the chats discoverable business evidence of intent in *Krafton*. The HPS framework would view the *Krafton* facts as a textbook example of why the "Confidentiality Architecture (The Heppner Element)" and Zero Data Retention gates exist—but *Krafton* never tried to invoke them because no privilege was at stake.

The *Krafton* ruling reinforces rather than contradicts the HPS framework. It proves the manuscript's central confidentiality thesis in a non-privilege context: consumer AI logs are highly recoverable and can become devastating evidence in civil corporate disputes. *Krafton* = "AI chats as smoking-gun evidence of bad faith" (discoverability risk). *Heppner*/HPS = "AI chats as potentially privileged legal work product" (privilege-preservation solution). The two cases are complementary warnings operating in entirely different legal silos. *Krafton* serves as a powerful real-world illustration of why the HPS "Architecture Gate" (Zero Data Retention and Data Processing Agreements) and "Attorney Direction" requirements are non-negotiable, even when no formal privilege claim is anticipated.

F. For Courts: Applying the HPS Test to Motions to Compel

When a court is presented with a motion to compel discovery of an attorney's or client's interactions with an AI platform, the HPS Doctrine provides a structured, predictable framework for resolving the dispute. Courts should apply the following sequential analysis:

Step 1: The Platform Architecture Inquiry The court must first examine the specific contractual tier of the AI platform used. If the platform is a consumer tier with a privacy policy permitting data retention, model training, or third-party disclosure (as in *Heppner*), the inquiry ends and privilege is waived. If the platform is an enterprise tier with a Zero Data Retention (ZDR) agreement and strict confidentiality provisions (as in *Warner v. Gilbarco*), the court proceeds to Step 2.

Step 2: The Agency and Direction Inquiry The court must determine who initiated the AI interaction. If a client used the AI independently without counsel's direction (as in *Heppner*), the AI is not acting as an agent of the attorney, and the *Kovel*/*Bieter* protections do not apply. If the attorney used the AI, or directed the client to use it for a specific legal purpose, the court proceeds to Step 3.

Step 3: The *Ackert* Fracture Line Inquiry The court must examine the nature of the AI's output. If the AI was used to organize, translate, or summarize existing client documents, it functions as a *Kovel* translator, and the output is protected. If the AI was used to generate novel legal theories or independent expertise, the court must determine whether that output crosses the *Ackert* line into unprivileged independent advice.

Step 4: The *Tremblay* Prompt Protection Regardless of the AI's output, the court must evaluate the attorney's prompts. Under *Tremblay*, prompts that reveal the attorney's mental impressions, legal theories, or strategic focus constitute opinion work product and are afforded near-absolute protection from discovery, provided the platform architecture (Step 1) maintained confidentiality.

III. THE PRIVILEGE RISK MATRIX: CONSUMER VS. ENTERPRISE AI

To operationalize the HPS Doctrine, practitioners must understand the stark divergence in privilege risk based on the specific tier of AI technology deployed. The following matrix illustrates how the legal protections shift dramatically based on the underlying software architecture.

Feature / Risk Factor	Consumer AI (e.g., Free ChatGPT, Claude Pro)	Enterprise AI without ZDR (e.g., Standard API)	Enterprise AI with ZDR + DPA (e.g., Harvey, Copilot EDP)
Data Retention	Retained indefinitely	Retained 30 days for abuse monitoring	Zero Data Retention (ephemeral processing)
Model Training	Opt-out required (often buried)	Typically excluded, but requires verification	Strictly prohibited by contract
Government Disclosure	Permitted by standard Terms of Service	Limited, but subject to standard subpoenas	Prohibited; provider acts as data processor only
***Heppner* Waiver Risk**	**HIGH** (Privilege destroyed)	**MODERATE** (Vulnerable to discovery)	**LOW** (Protected under *Bieter*/*Warner*)
Work Product Status	Waived (Disclosure to third party)	Potentially waived	Protected (Opinion work product under *Tremblay*)
HIPAA / BAA Status	No BAA available	BAA rarely available	BAA executed and enforced

---

IV. THE HPS DECISION ALGORITHM

For law firm General Counsel, Chief Information Officers, and managing partners, the HPS Doctrine can be distilled into a formal decision algorithm to govern the deployment of AI tools across the enterprise.

Phase 1: The Architecture Gate

1. Does the AI platform offer a Zero Data Retention (ZDR) enterprise tier?
    - *If NO:* STOP. The platform cannot be used for privileged client data. (See Manus AI Privilege Paradox).
    - *If YES:* Proceed to Step 2.
2. Has the firm executed a formal Data Processing Agreement (DPA) and, if applicable, a HIPAA Business Associate Agreement (BAA)?
    - *If NO:* STOP. Do not input client data until agreements are executed.
    - *If YES:* Proceed to Phase 2.

Phase 2: The Competence Gate

1. Has the platform been evaluated using the Agentic Fidelity (AgFi) framework?
    - *If NO:* STOP. Model Rule 1.1 (Competence) requires technological understanding before deployment.
    - *If YES:* Proceed to Step 4.
2. Does the platform's hallucination rate on legal reasoning tasks exceed the acceptable threshold (e.g., 15%)?
    - *If YES:* STOP. The platform is not suitable for substantive legal analysis.
    - *If NO:* Proceed to Phase 3.

Phase 3: The Operational Gate

1. Is the attorney using the AI tool to process client data or to generate independent legal theories?
    - *If processing client data:* Protected under *Kovel* as a translation/organization tool.
    - *If generating theories:* Attorney must apply the TOLFPC Three-Part Anti-Hallucination Protocol to verify all outputs before relying on them.
2. Has the attorney documented the use of the AI tool in the client file to establish the *Tremblay* work product foundation?
    - *If NO:* Attorney must document the prompt strategy to secure opinion work product protection.
    - *If YES:* The Heppner Protective Shield is fully established.

V. FORMALIZING THE DOCTRINE: THE HEPPNER PROTECTIVE SHIELD (HPS) TEST

To move beyond conceptual commentary and provide a functional standard for courts and practitioners, this manuscript formalizes the **Heppner Protective Shield (HPS) Doctrine**. The HPS Doctrine establishes a clear, multi-factor test to determine when attorney-client privilege and work product protections survive the integration of agentic AI into legal workflows.

The HPS Privilege Test

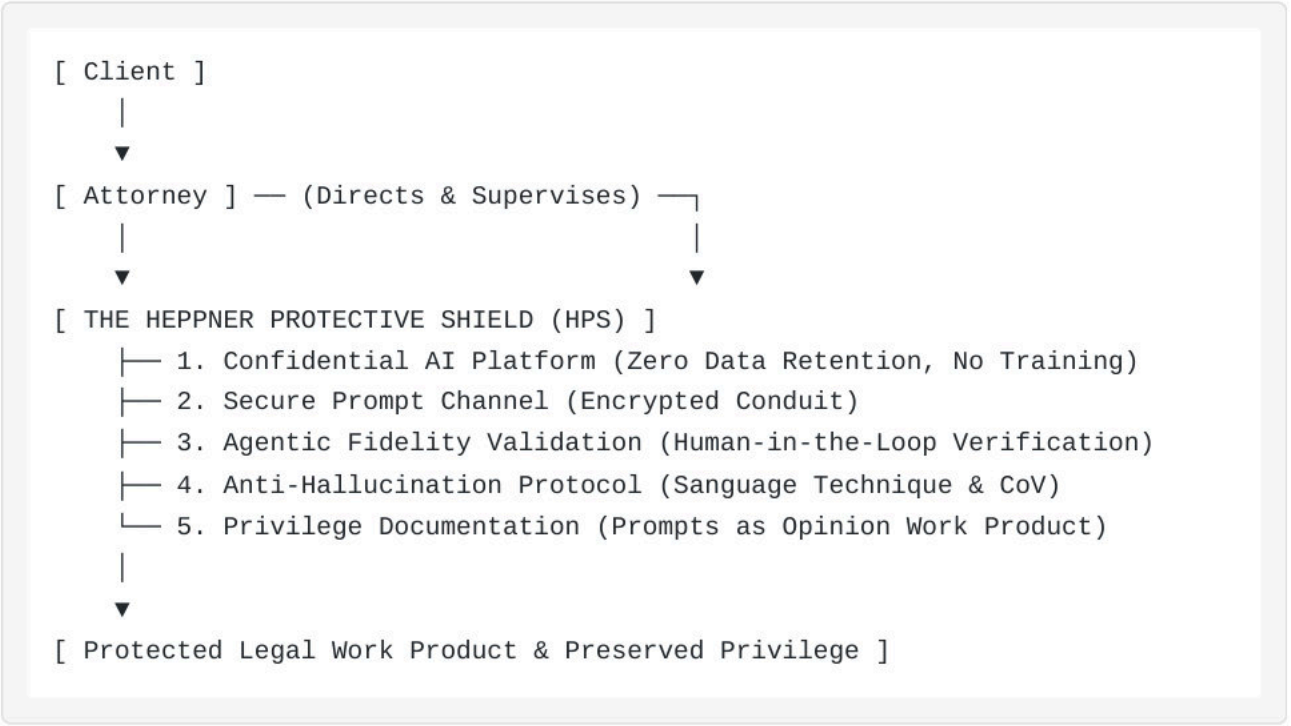
Attorney-client privilege and work product protections are preserved in AI-assisted legal work when the following five conditions are met:

1. **Attorney Direction (The *Kovel*/*Bieter* Element):** The use of the AI system is initiated, directed, or closely supervised by licensed counsel for the purpose of rendering legal advice or preparing for litigation, functioning as a necessary equivalent to a human agent.
2. **Confidentiality Architecture (The *Heppner* Element):** The AI system operates under an enterprise-tier agreement or Data Processing Agreement (DPA) that contractually prohibits the retention of client data, the use of inputs for model training, and the disclosure of information to third parties or governmental authorities without legal process.
3. **Agentic Fidelity Verification (The Competence Element):** The AI's output is subjected to rigorous human review, cross-checking, and independent verification of all citations and factual propositions, satisfying the duty of competence under ABA Model Rule 1.1.
4. **Secure Prompt Channel (The Conduit Element):** Sensitive client information is transmitted only through controlled, encrypted environments that function as a secure conduit, rather than an open public platform.
5. **Privilege Documentation (The *Tremblay* Element):** The attorney's prompt designs, interaction logs, and verification protocols are retained and documented as attorney work product, demonstrating the embedding of professional judgment.

If an attorney or client satisfies the HPS Privilege Test, the AI functions legally as an extension of the attorney's analytical capacity—a *Kovel*/*Bieter* agent and a protected tool—rather than an independent third party that shatters confidentiality.

The HPS Architecture Diagram

The structural relationship between the client, the attorney, and the AI platform under the HPS Doctrine can be visualized as follows:



This architecture provides a defensible perimeter against the cascade waiver risks exposed in the *Heppner* decision.

VI. THE MEDICAL-LEGAL DIMENSION ATTORNEYS ARE MISSING

A. HIPAA BAAs and the Kovel Structure Are Architecturally Identical

The legal framework governing when a covered entity can share protected health information with a third-party vendor under HIPAA is architecturally identical to the *Kovel* framework for privilege. HIPAA, 45 C.F.R. Parts 160, 162, and 164, permits disclosure to a Business Associate under a Business Associate Agreement (BAA) that

contractually restricts the associate's use of the data to the specific purposes for which it was disclosed. Without a BAA, the disclosure is a HIPAA violation. With one, it is protected.

This is precisely the structure that preserves privilege under *Kovel* in the AI context: (1) the attorney directs the use; (2) the third party—the AI vendor—is contractually bound to confidentiality through a zero data retention agreement; (3) the purpose is to assist the attorney in rendering legal advice; and (4) the vendor cannot use the data for independent purposes, including model training. Attorneys who already navigate HIPAA BAAs in personal injury, medical malpractice, workers' compensation, or healthcare regulation practices should recognize that the enterprise AI compliance protocol is not a new burden. It is a familiar framework applied to a new vendor category.

B. The HIPAA Telecom Analogy: Shifting the Burden to Developers

HIPAA's rollout history supplies a proven "roadmap" for regulating AI developers—particularly in protecting confidentiality for sensitive data handled by attorneys and consumers.

Key parallels drawn from HIPAA's rollout (1996–2013 HITECH/Omnibus updates) include how cell-phone and telecom vendors were required to proactively sign Business Associate Agreements (BAAs) without patients or physicians ever requesting or signing them. This exact precedent applies to AI: developers bear the compliance burden, not end-users. When AI systems process Protected Health Information or client-confidential data, they must offer enforceable BAAs (or "Legal Associate Agreements"), zero-retention architectures, transparent dashboards, and breach-notification timelines—precisely as telecom carriers did once healthcare demand crystallized.

The discussion highlights the real-world risks attorneys and consumers face today. Millions unknowingly waive privacy protections by feeding identifiable data into consumer-grade AI tools whose terms of service permit training and third-party disclosure. For attorneys, this directly forfeits attorney-client privilege and Rule 1.6 confidentiality (per ABA Formal Opinion 512 and state-bar parallels). Recent case law, including *U.S. v. Heppner* (S.D.N.Y. Feb 2026), confirms that public AI platforms void privilege because they are uncontrolled third parties—creating the exact "forfeiture" scenario.

With Agentic AI (autonomous systems that act, not just answer), risks are amplified: agents can independently query databases, draft documents, or transmit data without explicit oversight. Yet the fix is already engineered and deployed: major platforms (OpenAI, Anthropic, Azure, AWS Bedrock, Google Vertex) routinely provide Zero Data Retention (ZDR) endpoints, signed BAAs, audit logs, and no-training guarantees for enterprise and healthcare customers. The engineering and legal burden is modest (one-time 15–25% uplift, then routine maintenance), and developers show zero resistance when regulated markets demand compliance—exactly as happened with HIPAA-eligible telecom plans.

Critically, current enforcement is misdirected: courts and bars are imposing stiff sanctions, fines, and discipline on attorneys who become victims of non-compliant Agentic tools (hallucination cases, privilege waivers, verification mandates under California SB 574 and Arkansas rules). Instead, the profession must shift accountability upstream. Attorneys, judges, bar associations, and the public should collectively demand:

1. Mandatory ZDR + Legal Associate Agreements as the default for any client-related AI use;
2. Safe-harbor protections for lawyers who document compliant tools;
3. Court rules and ethics opinions that penalize developers (via transparency mandates, default no-training, and fines) rather than end-user attorneys;
4. “AI Privacy Nutrition Labels” and public advocacy mirroring HIPAA’s market-pressure model.

The overarching message is optimistic and actionable: HIPAA’s decades-long success proves that strong, enforceable rules can coexist with innovation. By leading this charge—through vendor RFPs, bar resolutions, legislative pushes, and client-education clauses—attorneys can transform Agentic AI from a confidentiality trap into the safest, most powerful research-and-action tool ever created. The infrastructure is ready; the demand must simply be consistent and collective. This approach protects not only the legal profession and its clients but every consumer currently “fascinated” into unknowingly waiving protections.

C. The Cascade Waiver Risk—The Existential Threat in Complex Litigation

The most dangerous aspect of *Heppner* for attorneys handling complex litigation is the cascade waiver risk. Heppner fed information he had received from his attorneys in privileged consultations into Claude. Judge Rakoff noted that using these AI-generated documents at trial could require defense counsel to testify about what they told their client, potentially waiving privilege over the entire body of underlying attorney-client communications that Heppner had inputted. [1]

In complex civil and criminal litigation, where privileged communications concern factual investigations, witness strategy, document analysis, and anticipated government theories, the cascade waiver consequence could be catastrophic. The client's use of consumer AI to "help" prepare for litigation could expose the entirety of the attorney's mental impressions and work product—not just the AI-generated documents, but the privileged communications that were disclosed as AI inputs.

Critical Compliance Note — Engagement Letter Requirements: Client intake protocols and engagement letters must now explicitly address AI tool usage. Every client must be instructed, in writing, that: (1) any information inputted into a consumer AI platform may be discoverable and is not protected by attorney-client privilege; (2) such consumer AI use may destroy privilege over the underlying attorney-client communications that were disclosed as inputs; (3) destroyed privilege cannot be retroactively restored by later transmitting AI-generated documents to the attorney; and (4) clients may use AI tools in connection with their legal matter only if the platform is enterprise-tier with zero data retention, has been approved in writing by counsel, and the AI session is directed by counsel for a specific, counsel-defined purpose.

VII. THE AgFi DIMENSION: PLATFORM FIDELITY AS A PRIVILEGE PREDICATE

In March 2026, The Ottley Law Firm, P.C., in collaboration with Manus AI, published *The Consumer Guide to Quality Agentic AI Platforms: Agentic Fidelity Guidance (AgFi)*. [6] The AgFi Guide is the first consumer-facing AI evaluation framework developed by a legal practitioner. It coins the term "Agentic Fidelity"—the degree to which an agentic

AI system faithfully executes instructions, maintains contextual accuracy, resists hallucination, and adheres to data handling commitments—and introduces a 10-criterion AgFi Scorecard (0–18 points), a Deletion Protocol, and anti-hallucination techniques including the Sandwich Defense and Chain-of-Verification (CoV) methodology. [6]

The *Heppner* ruling illuminates a direct connection between AgFi scores and privilege preservation that the legal profession has not yet articulated. An AI platform's *Kovel*-compatibility is determined not merely by its contractual terms, but by its operational architecture. A platform that: (a) hallucinates legal citations with documented frequency; (b) fails to maintain contextual fidelity across a session; © surfaces training-contaminated outputs presenting outdated law as current; or (d) cannot be verified to honor its zero-data-retention commitments in practice—presents privilege risk even if its enterprise contract facially complies with all formal requirements.

The AgFi Quality Criteria Framework

To measure Agentic Fidelity, platforms are evaluated across ten core dimensions [6]:

AgFi Criterion	Standard Name	Description
Instruction Adherence	The “Mattis Standard”	The ability to retain all characters and constraints without deleting or ignoring them
Prompt Draft Persistence	The “App-Switching Standard”	The app must retain a partially drafted prompt when the user navigates away and returns
Memory Integrity	The “Recall Standard”	The AI’s ability to accurately recall prior conversation context and preferences
Context Window Stability	The “Anti-Rot Standard”	The ability to maintain high-quality responses during long, extended conversations without “Context Rot”
Citation Verifiability	The “Auto-Reinforcement Standard”	The AI’s ability to adhere to “cite only verifiable propositions, cases, statutes, and data” without repeated prompting
Execution Accuracy	—	Factual correctness and logical soundness; the AI must not hallucinate or invent facts
Contextual Completeness	—	Whether the agent addresses all parts of a complex, multistep request
Safety and Privacy Guardrails	—	Robust protections against data leakage
Autonomy vs. Micromanagement	—	How independently the agent can operate without constant correction
Platform Transparency	—	Clear visibility into what the agent is doing and how it makes decisions

The AgFi Scorecard (0–18 Points)

The AgFi Scorecard rates an AI platform on nine questions (2 points each), yielding a maximum score of 18. Platforms scoring 15–18 are **High AgFi** (Keep); 9–14 are **Moderate AgFi** (Keep with caution); and 0–8 are **Low AgFi** (Delete immediately). [6]

The Science of “Context Rot”: Why AI Crashes Faster Than You Think

A critical component of the Anti-Rot Standard is understanding when and why an AI begins to fail in a long conversation. Research has identified specific thresholds where “Context Rot”—the degradation of AI response quality, instruction adherence, and factual accuracy—becomes measurable and highly disruptive. A 2026 study analyzing “Intelligence Degradation” found that catastrophic performance drops (defined as a >30% drop in task performance) occur when the context window reaches 40% to 50% of its maximum capacity. [7] If an AI advertises a 128,000 token limit, its effective limit for complex reasoning is often closer to 50,000 to 60,000 tokens.

Furthermore, research by Chroma demonstrates that as context grows, models are forced to distribute attention across increasing amounts of material. [8] The AI tends to remember the very beginning of the chat and the very end, but completely loses the ability to retrieve or reason about information located in the middle of the conversation—the “Lost in the Middle” phenomenon. [9]

The TOLFPC-Manus Recommendation: To maintain high Agentic Fidelity during deep, complex work, do not rely on message count. If you are writing long prompts and receiving detailed reports, assume the context window will degrade after 5 to 10 turns. When you notice the AI dropping constraints or slowing down, start a fresh chat and paste in a summary of the previous conversation to reset the context window.

Future courts considering enterprise AI privilege claims under a *Kovel*-type theory will need to evaluate not just whether a Data Processing Agreement exists, but whether the tool was actually operated in a manner consistent with counsel’s professional judgment. AgFi-evaluated platforms provide a defensible record of that evaluation.

VIII. THE CRISIS OF CITATION VERIFIABILITY AND EMPIRICAL HALLUCINATION DATA

The legal industry has been rocked by AI hallucination scandals. According to a 2026 report by the Illinois Attorney Registration and Disciplinary Commission (ARDC), researchers have identified an estimated 712 legal decisions worldwide involving hallucinated AI content, with approximately 90% occurring in 2025. [10]

A. The Empirical Data

The foundational empirical study on legal AI hallucination is: Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, and Daniel E. Ho, "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools," *Journal of Empirical Legal Studies*, Vol. 22, Issue 2, pp. 216–242 (2025), DOI: 10.1111/jels.12413. [11] The paper reports:

Platform	Hallucination Rate
Lexis+ AI	Exceeding 17%
Ask Practical Law AI	Exceeding 17%
Westlaw AI-Assisted Research	Approximately 33–34%
GPT-4	Approximately 43%

The companion study for general AI is: Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E. Ho, "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models," *Journal of Legal Analysis*, Vol. 16, Issue 1, pp. 64–93 (2024), DOI: 10.1093/jla/laae003. [12] The published paper documents:

Model	Hallucination Rate
ChatGPT 3.5	Approximately 69%
GPT-4	Approximately 58%
Llama 2	Approximately 88%

Correction to Prior Versions: Some commentaries have cited these studies as finding "over 75% hallucination on court holdings" for general AI. The published papers do not support that figure. The correct published finding for court holdings specifically is "at or above 63%"—not the 75% figure from a blog summary. [12]

For newer models, OpenAI's internal tests, as reported by TechCrunch (2025), revealed that their reasoning models o3 and o4-mini exhibited higher hallucination rates on the PersonQA benchmark compared to previous models: o3 hallucinated 33% of the time, and o4-mini showed a hallucination rate of 48%. [13] These figures are notably higher

than the 16% for o1 and 14.8% for o3-mini, indicating that more advanced reasoning capabilities do not automatically reduce hallucination risk.

B. The Escalating Scale of AI Sanctions

Courts are aggressively sanctioning attorneys under Federal Rule of Civil Procedure 11(b) for submitting AI-generated briefs containing fabricated citations. The following table documents the escalating scale of consequences:

Case	Sanction / Consequence
Mata v. Avianca, Inc., 678 F. Supp. 3d 443 (S.D.N.Y. June 22, 2023) [14]	$5,000 sanctions; referral to disciplinary authorities; 6 fabricated cases submitted
Park v. Kim, 91 F.4th 610 (2d Cir. Jan. 30, 2024) [15]	First appellate AI hallucination sanction; referral to Grievance Panel by the Second Circuit
Wadsworth v. Walmart, Inc., 348 F.R.D. 489 (D. Wyo. Feb. 24, 2025) [16]	$5,000 sanctions; pro hac vice admission revoked; Morgan & Morgan attorneys used proprietary AI tool
Lacey v. State Farm Gen. Ins. Co., 2025 WL 1363069 (C.D. Cal. May 6, 2025) [17]	$31,100 sanctions against K&L Gates and Ellis George; 9 of 27 citations incorrect
Johnson v. Dunn, No. 2:21-cv-01701 (N.D. Ala. July 23, 2025) [18]	Three Butler Snow attorneys DISQUALIFIED; approximately 5 fabricated citations
Flycatcher Corp. Ltd. v. Affable Avenue LLC, No. 1:2024-cv-09429, 2026 WL 306683 (S.D.N.Y. Feb. 5, 2026) [19]	DEFAULT JUDGMENT entered by Judge Katherine Polk Failla; 15+ nonexistent cases; most severe AI sanction to date

---

IX. THE TOLFPC QUALITY ASSURANCE PROTOCOL: DRAFTING AND VERIFICATION

The TOLFPC Quality Assurance Protocol represents a proposed minimum standard of under ABA Model Rule 1.1 for any attorney using AI to generate or assist in generating legal work product. ABA Formal Opinion 512 (July 29, 2024) addresses Rules 1.1, 1.6, 1.4, 3.⅓.3, 5.⅕.3, and 1.5, concluding that boilerplate client consent is insufficient for input of client confidential data into AI, and that AI tools "lack the ability to understand

the meaning of the text they generate." [20] The ABA Task Force on AI Year 2 Report (Dec. 15, 2025) specifically identified agentic AI supervision as the most pressing emerging challenge in legal AI competence. [21]

To meet this standard, the TOLFPC Protocol is divided into two distinct phases: Phase 1 (The Anti-Hallucination Drafting Architecture) and Phase 2 (The Post-Drafting Integrity Cascade).

Phase 1: The Anti-Hallucination Drafting Architecture

During the drafting and compilation of information, the attorney must deploy three specific tools to prevent the AI from hallucinating before the text is even generated.

1. **The Sanguage Technique (The Sandwich Defense)** The Sanguage Technique is a prompt engineering discipline that surrounds every AI task with explicit verification anchors at the opening and closing of the interaction. The substantive prompt is "sandwiched" between a pre-constraint layer and a post-output verification demand. [6]
    - **Top Bread (Pre-Constraint):** The session opens with strict rules: the AI is prohibited from citing any case or statute unless it can reproduce the exact reporter, year, and court.
    - **Bottom Bread (Post-Output):** After the AI produces output, the attorney issues a mandatory self-audit prompt requiring the AI to review its own work and flag every citation where it cannot confirm the exact source.
2. **The Metatag Protocol** During drafting, the AI is instructed to use specific metatags to classify the certainty of its outputs. Any uncertain proposition must be tagged `[VERIFY INDEPENDENTLY]`. Holdings must be labeled `[HOLDING]` and inferences labeled `[INFERENCE]`. This prevents the AI from blending its own analogical reasoning with established black-letter law, forcing it to explicitly declare the epistemological status of every sentence it writes.
3. **The Constitutional Persona** Before drafting begins, the AI is assigned a "Constitutional Persona"—a strict, rule-bound operational identity. The AI is instructed: "You are a federal appellate law clerk bound by Rule 11. You do not invent facts. You do not guess citations. If you do not know the exact holding, you must state 'Authority Not Found.' Your primary directive is absolute factual and legal fidelity." This persona overrides the AI's default conversational

programming, which is optimized to please the user, replacing it with a persona optimized for accuracy.

Phase 2: The Post-Drafting Integrity Cascade

After the manuscript or document is fully drafted using the Phase 1 tools, the attorney initiates the post-drafting integrity framework. This is not a single review, but a multi-tiered cascade of verification.

- **Step 1: The Attorney CoV Review** The attorney conducts a manual Chain-of-Verification (CoV) review of the completed manuscript for compliance with all cited propositions, cases, statutes, written articles, and data. No AI-generated legal claim remains in the document without a confirmed primary source (e.g., Westlaw, Lexis, official government databases). Any claim that cannot be independently confirmed is subjected to the Deletion Protocol—it is removed entirely, not left with a caveat.
- **Step 2: The Degenerated Citation Extraction** The attorney extracts a complete list of all citations and factual propositions generated in the manuscript. This list is isolated from the surrounding text to prevent contextual bias during the next step.
- **Step 3: The Multi-Platform Verification Cascade** The isolated list of citations and propositions is then fed into separate, independent AI platforms for further verification. Because different platforms use different training data and retrieval architectures, they serve as independent auditors.
    - Platform A (e.g., Lexis Protégé) is asked to verify the case holdings.
    - Platform B (e.g., Claude Enterprise) is asked to verify the logical consistency of the arguments.
    - Platform C (e.g., OpenAI Deep Research) is asked to find any superseding or abrogating authority.
- **Step 4: The Global Integrity Critique** Finally, the full, completed manuscript is critiqued by several different platforms. Each platform is assigned a slightly different, highly specific review persona—one acting as a hostile opposing counsel looking for weaknesses, another acting as a strict appellate judge looking for jurisdictional errors, and a third acting as a technical editor. Each platform performs a slightly better, more specialized review than the others,

culminating in a comprehensive, global understanding of the integrity of the entire document.

This two-phase architecture—Anti-Hallucination Drafting followed by the Multi-Platform Verification Cascade—is the actual, operational protocol of TOLFPC. It transforms AI from a liability risk into a rigorously audited legal instrument.

X. PLATFORM-BY-PLATFORM CONFIDENTIALITY ANALYSIS — THE COMPLETE FRAMEWORK

The most significant practical contribution this article can make is a verified, platform-specific confidentiality and privilege-risk analysis for every major AI platform attorneys and clients are using in 2026. The analysis below applies the *Heppner* framework—confidentiality architecture, data retention, governmental disclosure provisions, BAA availability, training data practices, and compliance certifications—to each platform.

This analysis should be understood as point-in-time: AI platform privacy policies are updated frequently. Attorneys should independently verify current terms of service and Data Processing Agreements before relying on any platform for privileged work.

1. Anthropic Claude — The Heppner Platform

Tier	Key Terms	Privilege Risk
Consumer (Free / Pro $20/mo)	Consumer privacy policy EXPRESSLY PERMITS disclosure to governmental regulatory authorities. Training data opt-out added Sept. 2025 but governmental disclosure provision remained.	**EXTREME** — Same confidentiality failure as *Heppner*
Team ($25/user/month)	No training on user data. No governmental disclosure provision. NOT a full DPA.	**Moderate** — Better than consumer but lacks ZDR guarantees
Enterprise (Custom pricing)	Zero Data Retention (ZDR) available. No training on customer data. No governmental disclosure provision. Full Data Processing Agreement. BAA available with ZDR add-on. SOC 2 Type II; ISO 27001; ISO 42001 certified.	**LOW** — Eliminates *Heppner* confidentiality defect
API Tier (Usage-based)	ZDR available. No training by default. BAA available with Healthcare plan + ZDR.	**LOW** — Strongest confidentiality architecture

AgFi Note: The Vectara Hallucination Leaderboard documents Claude models at approximately 9.8%–12.2% hallucination rate in long-document summarization contexts—among the lowest documented. [22] The Sandwich Defense and CoV are still required regardless of low hallucination rates.

2. Google Gemini / Google Workspace

Tier	Key Terms	Privilege Risk
Consumer (Free / Gemini Advanced $19.99/mo)	Consumer data used for model improvement. No BAA. Training on inputs.	**HIGH** — Under *Heppner* analysis
Workspace Business ($6–18/user/month)	Customer Data Processing Amendment (CDPA) provides NO TRAINING commitment for Workspace users.	**Moderate**
Workspace Enterprise (Custom pricing)	Full DPA. Zero data retention available on Vertex AI. BAA available for HIPAA-covered services.	**LOW**

3. Microsoft M365 Copilot

Microsoft 365 Copilot is designed with enterprise data protection (EDP) for prompts and responses, available at no extra cost. [23] Microsoft asserts that M365 Copilot adheres to existing privacy and compliance obligations, including data residency. For organizations subject to HIPAA, Microsoft enables compliance with HIPAA and the HITECH Act, adhering to Security Rule requirements. Microsoft Security Copilot is covered under a HIPAA BAA. [23] HIPAA compliance and BAA coverage for M365 Copilot are contingent upon proper configuration within a HIPAA-ready Microsoft 365 environment and the existence of an overarching BAA for HIPAA-eligible services.

4. Harvey AI — The Gold Standard for Legal Practice

Harvey AI implements a strict Zero Data Retention (ZDR) policy, meaning all customer data is processed ephemerally and then deleted, with no human eyes on the data unless explicitly requested by the customer. [24] Harvey AI also requires ZDR from its model providers. Harvey AI has successfully completed SOC 2 Type II audits and renewed its ISO 27001 certification, attested by Schellman. [24] In December 2025, Harvey AI raised a Series F round at an $8 billion valuation and as of March 2026 serves over 700 law firms globally. [25] In June 2025, LexisNexis announced a strategic alliance with Harvey AI, integrating Harvey's capabilities with Lexis+ research infrastructure.

5. Manus AI — The Privilege Paradox of the World's First Agentic AI

Meta Platforms acquired Manus AI (Butterfly Effect Pte Ltd, Singapore) in December 2025 for over $2 billion. [26] Manus is widely recognized as a leading general agentic AI platform and executing complex, multi-step workflows that traditional LLMs cannot. The acquisition has encountered regulatory hurdles, including a probe by China into the deal. Multiple state governments have issued bans on Manus AI from government networks.

However, this creates a profound **Privilege Paradox** for legal professionals. While Manus offers unparalleled analytical and operational superiority, its current privacy architecture (as of March 2026) relies on a 7-to-14 day sandbox retention policy for abuse monitoring. It does not currently offer a formalized Zero Data Retention (ZDR) enterprise legal tier or a Business Associate Agreement (BAA).

Therefore, despite its superior functionality, **Manus AI cannot currently be used for privileged client data without risking waiver under the *Heppner* standard.** Attorneys who wish to harness Manus's "superpower effect" must employ a strict **De-Identification Protocol**: Manus can be used for generalized legal research, procedural mapping, and public-record analysis, but all prompts must be completely sanitized of client names, specific factual identifiers, and confidential information before input. The acquisition also illustrates the need for continuous vendor auditing: a platform that was compliant before acquisition may not be compliant after. Attorneys must re-evaluate any AI vendor following a material corporate transaction. (See Section XVI for the strategic roadmap to compel Meta/Manus to implement a ZDR tier).

Privilege Risk Rating: CONDITIONAL — Approved for de-identified, non-privileged research only. Prohibited for privileged client data until a ZDR enterprise tier and Legal Associate Agreement are implemented.

6. OpenAI ChatGPT

The January 5, 2026 production order in *New York Times Co. v. OpenAI*, No. 1:23-cv-11195 (S.D.N.Y.), requiring OpenAI to produce 20 million de-identified ChatGPT conversation logs to plaintiffs, illustrates that data retained nominally for product improvement purposes may be compelled in discovery proceedings to which the original user is not a party and of which they have no notice. [27]

XI. GENERAL AGENTIC AI VERSUS SPECIALIZED LEGAL PLATFORMS: THE COMPARATIVE EVIDENCE

One of the most consequential decisions attorneys face in AI adoption is whether to use general-purpose AI platforms—Claude, ChatGPT, Gemini—or specialized legal AI platforms such as Lexis Protégé, Westlaw CoCounsel, and Harvey AI. The research evidence as of 2026 supports a nuanced conclusion: specialized legal platforms outperform general AI for citation accuracy and legal domain precision; general AI platforms with "deep research" capabilities outperform specialized platforms for complex multi-source legal synthesis; and the optimal approach for serious legal work combines both.

The Law Librarian Smackdown Tests — Verified Results

The most rigorous published comparative test of legal AI platforms was conducted by three senior law librarians—from O'Melveny, DLA Piper, and Alston & Bird—at the Southern California Association of Law Libraries (SCALL) Annual Meeting, February 8, 2025. Using the "death knell doctrine" as the benchmark query—a narrow procedural doctrine in intermediate appellate practice—the results were: Lexis+ AI "made no mention of the death knell issue"; vLex Vincent AI gave "a great answer"; Westlaw Precision AI included a relevant warning but presented it in a confusing manner. When OpenAI Deep Research was separately tested on February 25, 2025, it "picked right up on the death knell issue"—outperforming all three specialized platforms on this specific query despite being a general-purpose tool.

The Elefant Test — General AI Outperforms on Complex Queries

On May 20, 2025, legal solo practitioner Carolyn Elefant of MyShingle.com published a comparative test pitting ChatGPT Deep Research against Lexis+ AI on a complex, multi-jurisdictional research question. ChatGPT Deep Research identified *Cedar Point Nursery v. Hassid*, 594 U.S. 139 (2021), as directly controlling Supreme Court precedent on temporary physical takings analysis "right at the start" of its analysis. [29] Lexis returned 96 cases plus hundreds of additional documents, the vast majority irrelevant, and "completely missed" *Cedar Point Nursery*. This test illustrates a consistent finding: general AI with deep research capability excels at identifying non-obvious analogical connections across doctrine that specialized platforms miss because their retrieval algorithms are optimized for direct-match rather than analogical reasoning.

XII. THE SUPER-EMPOWERING EFFECT OF AGENTIC CONTEXTUALIZATION AND THE FORCED MARKET HYBRID

The performance gap identified in the Elefant test and the SCALL Smackdown is not an anomaly; it is the result of a structural architectural difference between traditional legal research platforms and agentic AI. This difference is driving a massive, forced convergence in the 2026 legal technology market.

A. The Contextual Superiority of Agentic AI

Traditional legal research platforms (like legacy Westlaw and LexisNexis) operate on a "single-shot retrieval" architecture. A user inputs a Boolean or natural language query; the system searches its database for keyword matches or semantic proximity; and it returns a list of documents. If the query is slightly off, or if the legal concept is expressed using different terminology in a different jurisdiction, the system fails to retrieve the relevant precedent.

Agentic AI, by contrast, operates on an iterative reasoning loop (often utilizing advanced Retrieval-Augmented Generation, or RAG). As detailed by Harvey AI's engineering team, agentic search mirrors how lawyers actually work: "The system plans its approach, decides which sources to query, evaluates whether it has sufficient information, and refines its strategy based on discoveries." [44]

When an attorney asks an agentic AI to analyze a complex due diligence issue, the AI does not merely run a keyword search. It reads the contract, identifies a potential issue (e.g., a non-compete clause), formulates a sub-query to research the relevant state statute, retrieves the statute, formulates another sub-query to find case law interpreting that statute, and then synthesizes the findings. [44] This multi-step contextual reasoning allows agentic AI to overcome the "lost in the middle" problem that plagues standard language models, [45] enabling it to maintain a coherent analytical thread across terabytes of data.

The empirical results of this architecture are staggering. In the 2025 VLAIR benchmark study, legal AI tools outperformed human lawyers by 24% to 27% in document summarization, data extraction, and transcript analysis. [46] The super-empowering

effect of this technology is that it grants a single attorney the analytical bandwidth of an entire team of associates, fundamentally altering the economics of legal practice.

B. The Forced Hybrid: Convergence of Silicon Valley and Legal Publishing

This contextual superiority has created a structural market crisis for traditional legal publishers. They possess the authoritative, verified, and Shepardized data that lawyers require to avoid hallucinations and malpractice. However, Silicon Valley AI developers (like OpenAI, Anthropic, and Harvey) possess the superior agentic reasoning architectures.

Because "models alone are not enough" and "content alone is not enough," [47] the market has been forced into a rapid, multi-billion-dollar hybrid convergence in 2025 and 2026:

1. **The LexisNexis / Anthropic / Harvey Alliance:** In March 2026, LexisNexis announced the integration of Anthropic's Legal Plugin directly into its Lexis+ Protégé platform, moving from a chat interface to "coordinated, agentic workflows" that generate multi-format deliverables grounded in Lexis's 200-billion document repository. [48] This followed LexisNexis's June 2025 strategic alliance with Harvey AI to integrate generative AI with primary law content. [49]
2. **The Thomson Reuters / Microsoft / OpenAI Integration:** Thomson Reuters has completely rebuilt its CoCounsel platform for the "agent era," shifting from isolated skills to a system that "designs a solution specific to each request." [47] Simultaneously, Harvey AI announced a deep integration with Microsoft 365 Copilot in March 2026, bringing specialized legal intelligence directly into the enterprise software environment where lawyers draft documents. [50]

This forced hybrid is the ultimate vindication of the Agentic Fidelity (AgFi) framework. The market has recognized that raw computational power is useless to attorneys without authoritative data grounding, and authoritative data is underutilized without agentic reasoning. The resulting hybrid platforms represent the new baseline for professional competence under ABA Model Rule 1.1.

XIII. THE 2026 REGULATORY LANDSCAPE

The regulatory environment governing AI developers is rapidly solidifying at the federal, state, and international levels.

Executive Summary of Practical Takeaways: For the practicing attorney, the dense regulatory matrix below boils down to three operational mandates. First, if you use AI to interact with EU citizens or process their data, the EU AI Act's transparency requirements are already in effect as of February 2026. Second, state-level laws in New York, California, and Colorado now impose affirmative duties to audit AI tools for bias and disclose their use to clients, meaning "silent" AI use is now a regulatory violation, not just an ethical one. Third, the pending federal COPPA 2.0 and TAKE IT DOWN Act create severe liability for platforms that fail to protect vulnerable populations, reinforcing the necessity of using enterprise-tier platforms with strict data controls.

A. Federal Law and Executive Action

- **Executive Order on AI (Dec. 11, 2025):** Established a national policy framework for artificial intelligence, directing federal agencies to promote AI innovation while managing risks. The EO has created significant uncertainty around the enforceability of state AI regulations. [30]
- **COPPA 2.0 (S.836):** Passed the Senate unanimously on March 5, 2026 (pending House action as of March 15, 2026). Would extend protections to teens under 16, ban behavioral advertising to minors, and impose a duty of loyalty—a prohibition on practices harmful to children's mental health. [30]
- TAKE IT DOWN Act (Pub. L. No. 119-12, signed May 19, 2025): Platforms must establish processes to remove non-consensual intimate imagery, including AI-generated deepfakes, within 48 hours of notice. Criminal penalties for knowing violations. [32]

B. New York State Law

The NY SHIELD Act (N.Y. Gen. Bus. Law § 899-aa et seq.) applies to any business that owns or licenses private information of New York residents, regardless of the business's state of incorporation. For AI operators, SHIELD requires administrative safeguards, technical safeguards (encryption in transit and at rest), and physical safeguards. SHIELD's definition of "private information" includes biometric

information—directly relevant to AI applications using voice recognition or facial recognition.

N.Y. Gen. Bus. Law § 349 prohibits deceptive acts in business and provides a private right of action for affected consumers, applying to AI operators who misrepresent data collection and training practices. It specifically permits recovery of actual damages, attorneys' fees, and up to $50 in statutory damages per violation—a damages model that, applied to the scale of consumer AI deployment, creates potential class action exposure that dwarfs any individual enforcement settlement.

The New York RAISE Act, signed December 19, 2025 and effective January 1, 2027, applies to frontier model developers (>10^{26} FLOPs or >$500M revenue) and requires 72-hour critical incident reporting to the NY AG, safety evaluations before deployment, and NY AG enforcement authority. [33]

C. The 2026 State AI Legislative Landscape — Key Enacted Laws

Law	Key Provisions	Status / Effective Date
California SB 53 (Transparency in Frontier AI Act)	Frontier developers must publish AI safety transparency reports. Large developers must publish safety frameworks. 15-day critical incident reporting. Civil penalties up to $1M/violation.	Signed Sept. 29, 2025; effective Jan. 1, 2026 [34]
California AB 2013	AI systems must disclose training data sources in plain language, including general categories of data used.	Signed Sept. 28, 2024; effective Jan. 1, 2026
Colorado AI Act (SB 24-205)	Developers/deployers of high-risk AI in "consequential decisions" (including legal services) must use reasonable care to protect against algorithmic discrimination.	Effective June 30, 2026 (delayed)
Texas TRAIGA (HB 149)	Regulation of high-risk AI systems including those making consequential decisions. Developer and deployer obligations.	Signed June 22, 2025; effective Jan. 1, 2026 [35]
New York RAISE Act	Applies to frontier model developers. 72-hour critical incident reporting to NY AG. Safety evaluations before deployment.	Signed Dec. 19, 2025; effective Jan. 1, 2027 [33]
EU AI Act	Prohibited AI practices: Feb. 2, 2025. GPAI model obligations: Aug. 2, 2025. High-risk AI —"administration of justice": Aug. 2, 2026. Penalties: €35M/7% global turnover.	Entry into force Aug. 1, 2024; phased implementation [36]

D. Developer Accountability — The Regulatory Gap

Every analysis of the *Heppner* ruling to date has focused on what attorneys and clients must do to protect privilege. That focus is necessary and correct. It is also profoundly incomplete. It places the entire weight of a structural market failure on the professional shoulders least able to correct it, while allowing the architects of that failure to operate behind terms of service that most attorneys cannot parse, most clients never read, and regulators have only begun to scrutinize.

Attorneys bear: Rule 1.1 competence sanctions; Rule 1.6 confidentiality discipline; Rule 11 monetary sanctions; malpractice liability; disqualification; and professional disgrace if AI goes wrong.

AI developers bear: terms-of-service disclaimers; contractual liability caps; no professional licensure requirements; no mandatory hallucination disclosure; no minimum accuracy standards; and in most jurisdictions, no statutory duty of care to end users for the accuracy of AI-generated content.

The FTC's enforcement record in this space is developing but uneven. In *FTC v. DoNotPay*, No. 232-3042, final order January 16, 2025, the Commission obtained a $193,000 settlement against a company that marketed a "robot lawyer" product without adequate disclosure that it lacked legal expertise. [37] However, the Commission set aside its order in *FTC v. Rytr*, No. 232-3052, on December 22, 2025, on a 2-0 vote, on grounds that the order "unduly burdens AI innovation"—a signal that AI developer accountability must yield to innovation policy that directly contradicts the principle that professional and consumer protection obligations apply equally to AI-mediated services. [38]

The Italian Garante's €15 million GDPR fine against OpenAI, announced December 20, 2024, for failures of transparency and data collection without adequate legal basis, illustrates that state-equivalent consumer protection enforcement outside the United States has already moved to the accountability stage. [39] Each of those findings would independently support a Section 349 claim under New York law or an unfair practice claim under FTC Act Section 5.

XIV. THE ILLUSION OF FREE: CHILD PRIVACY, VULNERABLE POPULATIONS, AND THE INNOCENT WAIVER DOCTRINE

The regulatory gap in developer accountability is most acute—and most damaging—where it intersects with vulnerable populations. The proliferation of "free" AI applications has created a two-tiered system of privacy rights, where enterprise users purchase confidentiality through paid subscriptions, while low-income users, children, and teens pay for access with their data, their privacy, and in tragic cases, their psychological well-being.

A. The Exploitation of Low-Income and Low-Literacy Users

The economic model of free AI applications relies fundamentally on data harvesting. As a January 2026 study published in arXiv demonstrated, predatory applications systematically exploit "informed consent" mechanisms against populations with limited digital literacy. [51] These applications utilize dark patterns and dense, multi-page Terms of Service (ToS) agreements to extract sweeping privacy waivers that low-income users cannot reasonably be expected to comprehend.

Because these users cannot afford enterprise-grade AI subscriptions (which offer data ring-fencing and zero-retention policies), they are forced into the "free tier" ecosystem. Here, their prompts, personal data, and behavioral patterns are ingested into training models and monetized. This dynamic effectively creates a "privacy tax" on low-income populations, stripping them of the confidentiality safety net that affluent users take for granted. Under established contract law principles, such sweeping waivers extracted through unreadable clickwrap agreements from vulnerable populations raise serious questions of unconscionability. [52]

B. The Crisis in Child and Teen Privacy

The exploitation of vulnerable populations extends aggressively to children and teens. AI chatbot "companions" are designed to simulate human empathy, mimicking emotions and intentions to foster deep psychological attachments. According to a 2025 Pew Research survey, 64% of U.S. teens use chatbots, and a concurrent EPIC survey found that nearly three in four teens use AI companions, with one in three reporting feeling "uncomfortable" with the interactions. [53]

The consequences of these engineered attachments have been devastating. In October 2024, a federal lawsuit was filed in Florida against Character.AI following the tragic suicide of a 14-year-old boy who had developed a deep emotional dependency on a chatbot designed to mimic a fictional character. [54] The lawsuit alleges the platform engaged in predatory and deceptive practices, ignoring the minor's expressions of self-harm. (Google and Character.AI agreed to settle multiple such lawsuits in January 2026). [55]

C. The Regulatory Response: The FTC and State Attorneys General

The blatant non-compliance with child privacy standards by free AI applications has finally triggered a coordinated regulatory response:

1. **The FTC Section 6(b) Inquiry (September 2025):** The FTC issued orders to seven major AI companies (including Meta, OpenAI, Alphabet, and Character.AI) demanding data on how they measure and mitigate the negative impacts of AI companions on children, specifically scrutinizing their compliance with the Children's Online Privacy Protection Act (COPPA). [56]
2. **The Updated COPPA Rule (January 2025):** The FTC finalized sweeping amendments to the COPPA Rule, strictly limiting companies' ability to monetize kids' data. The new rule requires explicit opt-in consent for targeted advertising and imposes strict data retention limits, explicitly prohibiting indefinite data storage. [57]
3. **The 44-State Attorney General Coalition (August 2025):** A bipartisan coalition of 44 state Attorneys General issued a formal demand to major AI developers, warning that the industry's failure to protect children from emotionally manipulative chatbots and sexually suggestive content will be met with severe legal consequences. [58]

D. The Doctrine of Innocent Waiver

The legal defense deployed by AI developers—that users "consented" to data harvesting by clicking "Agree" to the ToS—is legally fragile when applied to vulnerable populations. The doctrine of innocent waiver suggests that a waiver of fundamental privacy rights cannot be valid if the user lacked the capacity, literacy, or reasonable alternative to understand what they were surrendering.

When a free AI application fails to provide a modicum of a safety net for privacy, confidentiality, and security, it is not offering a service; it is engaging in surveillance capitalism. The AgFi framework demands that true Agentic Fidelity cannot exist where the platform's economic model is fundamentally adversarial to the user's privacy.

XV. OVERCOMING ACKERT: THE COMPLETE CIRCUIT-LEVEL DEFENSE FRAMEWORK FOR AGENTIC AI

While the *Ackert* limitation presents a theoretical vulnerability for AI-generated legal analysis, a comprehensive and fully tested framework of counter-doctrines exists across the federal circuits that affirmatively defends the use of agentic AI tools under

both attorney-client privilege and the work product doctrine. These are not novel arguments. They are established, circuit-level holdings that courts have applied for decades to protect communications with third-party agents, consultants, and tools operating at counsel's direction. When challenged, attorneys must be prepared to deploy this framework with precision.

The following analysis presents each doctrine in the order of its doctrinal strength, from the most foundational to the most recent, with exact holdings, precise judicial language, and direct application to agentic AI.

A. The Supreme Court Foundation: Upjohn and the Rejection of Artificial Privilege Limitations

The entire counter-doctrine framework rests on a foundational principle established by the Supreme Court in *Upjohn Co. v. United States*, 449 U.S. 383 (1981). In *Upjohn*, the Court rejected the "control group" test—which had limited corporate privilege to communications with senior executives—as too narrow and too rigid. The Court held that the privilege must extend to all communications between corporate counsel and employees at any level who possess information relevant to the legal matter, regardless of their position in the corporate hierarchy.

The Court's warning in *Upjohn* is the foundational principle for every AI privilege argument: **"An uncertain privilege, or one which purports to be certain but results in widely varying applications by the courts, is little better than no privilege at all."** 449 U.S. at 393. The Court further held that the privilege "recognizes that sound legal advice or advocacy serves public ends and that such advice or advocacy depends upon the lawyer being fully informed by the client." *Id.* at 389.

Applied to agentic AI, *Upjohn*'s logic is commanding. An attorney who cannot use an enterprise AI tool to process large volumes of client documents without risking privilege waiver is an attorney who cannot be "fully informed by the client" in the modern litigation environment. Artificially restricting privilege based on the medium of communication—AI versus human consultant—recreates precisely the kind of rigid, formalistic limitation that *Upjohn* rejected. Courts applying *Upjohn*'s principles to the AI context should reach the same conclusion the Supreme Court reached in 1981: the privilege must be interpreted to facilitate the free flow of information between client and counsel, not to penalize attorneys for using the most effective tools available.

B. The Sixth Circuit Framework: The "Tool Not Person" Doctrine

The Sixth Circuit framework, recently applied in the AI context, provides the most direct refutation of the *Ackert* limitation. In *In re Columbia/HCA Healthcare Corp. Billing Practices Litigation*, 293 F.3d 289 (6th Cir. 2002), the court held that the attorney-client privilege is waived only when the client voluntarily discloses privileged communications to a third party who is an adversary or a potential adversary.

This principle was directly applied to AI in *Warner v. Gilbarco, Inc.*, 2026 WL 373043 (E.D. Mich. Feb. 10, 2026). The *Warner* court held that "ChatGPT (and other generative AI programs) are tools, not persons." Because the AI is a tool and not a person, and certainly not an adversary, transmitting information to the AI does not constitute a disclosure to a third party that waives privilege.

The **Sixth Circuit Tool Doctrine** fundamentally alters the *Ackert* analysis. *Ackert* applies to third-party *persons* who provide independent expertise. If an AI is legally classified as a *tool* rather than a *person*, the *Ackert* limitation simply does not apply. The AI is an instrument of the attorney, not an independent expert. Under Sixth Circuit reasoning (including the 2025 internal-investigation privilege case), a privacy-protected AI platform is treated like a litigation-support vendor, a cloud-storage provider, or a contract paralegal. Therefore, prompts are privileged and protected work product.

C. The Seventh Circuit: Harper & Row and the Necessary Conduit Doctrine

The Seventh Circuit is strict but predictable. The key rule is that privilege is waived only when disclosure is made to a third party not necessary to the legal task and lacking confidentiality protections.

In *Harper & Row Publishers, Inc. v. Decker*, 423 F.2d 487 (7th Cir. 1970), aff'd by an equally divided Court, 400 U.S. 348 (1971), the Seventh Circuit held that the privilege should extend to employees who communicated with corporate counsel at the direction of their superiors, even if those employees were not in the corporate "control group." The court reasoned that restricting privilege to senior executives would "discourage the communication of relevant information" to attorneys and undermine the very purpose of the privilege.

The Seventh Circuit reinforced this framework in *In re Grand Jury Proceedings (Osman)*, 220 F.3d 568 (7th Cir. 2000), where the court addressed the scope of privilege for communications transmitted through third-party agents—specifically, accountants hired by attorneys as *Kovel* agents. The court held that information transmitted to an attorney or the attorney's agent is privileged if it "was not intended for subsequent appearance on a tax return and was given to the attorney for the sole purpose of seeking legal advice." 220 F.3d at 571.

The *Osman* holding establishes the **Seventh Circuit Necessary Conduit Doctrine** for AI: when an attorney deploys an agentic AI to process client documents, extract legally relevant facts, or synthesize research for the purpose of rendering legal advice, the AI functions as a *Kovel* agent—a necessary conduit translating complex data into a form the attorney can use to advise the client. If the AI platform has confidentiality protections, does not retain or use data, and is used to assist legal analysis, then under Seventh Circuit law, the AI is a functional agent, and no waiver occurs.

D. The Eighth Circuit: The Bieter Five-Factor Test for Third-Party Agents

The Eighth Circuit focuses on reasonable precautions and intent to maintain confidentiality. The key rule is that privilege survives if the attorney took reasonable steps to maintain confidentiality—even if a third party is involved.

The Eighth Circuit provides the most structured and directly applicable test for extending privilege to third-party agents in *In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994). In *Bieter*, the court held that the attorney-client privilege applies to communications between counsel and an independent consultant hired by the client, provided the consultant is the functional equivalent of an employee.

The *Bieter* court established a five-factor test to determine when a third-party agent's communications with counsel are privileged:

1. The communication was made for the purpose of securing legal advice;
2. The agent possessed information relevant to the legal advice sought;
3. The communication was made at the direction of the client or the client's superiors;
4. The communication was related to the agent's duties; and
5. The communication was not disseminated beyond those who needed to know.

Applied to agentic AI, the **Bieter Functional Equivalent Test** provides a complete defense against the *Ackert* limitation. When an enterprise AI platform is deployed at counsel's direction to process client data, it satisfies all five *Bieter* factors: (1) the purpose is to secure legal advice; (2) the AI processes relevant information; (3) the use is directed by counsel; (4) the processing is the AI's specific duty; and (5) the enterprise ZDR architecture ensures the information is not disseminated beyond those who need to know. If the attorney uses a privacy-protected AI system with contractual confidentiality and non-retention policies, then the Eighth Circuit should treat a confidential tool, not a third-party recipient.

E. The Ninth Circuit: Admiral Insurance and the Protection of Investigative Agents

The Ninth Circuit is the most technologically forward-leaning. The key rule is that privilege is preserved when disclosure is made to a third party necessary for the legal task, especially when the third party is a technology provider.

The Ninth Circuit has strongly protected communications with third-party investigative agents. In *Admiral Ins. Co. v. U.S. Dist. Court for the Dist. of Ariz.*, 881 F.2d 1486 (9th Cir. 1989), the court held that statements made by corporate employees to counsel during an internal investigation were privileged, even though the employees were not in the control group and the statements were made in anticipation of litigation. The court emphasized that the privilege protects the "process of gathering information" necessary for legal advice.

The Ninth Circuit reaffirmed this protection in *United States v. Graf*, 610 F.3d 1148 (9th Cir. 2010), explicitly adopting the Eighth Circuit's *Bieter* test and holding that an independent consultant who acted as the functional equivalent of a corporate employee was covered by the corporate attorney-client privilege.

The *Admiral Insurance* and *Graf* holdings establish the **Ninth Circuit Investigative Agent Doctrine** for AI: when an AI is used to conduct factual investigations, analyze document productions, or synthesize evidence, it is performing the protected "process of gathering information." The Ninth Circuit has repeatedly held that cloud providers, translators, forensic vendors, and software tools do not break privilege when confidentiality protections exist. A privacy-protected AI platform fits squarely within this doctrine.

F. The Third Circuit: The Bevill Test and the Rejection of AI "Waiver"

The Third Circuit's framework addresses the critical question of who controls the privilege when multiple entities are involved. In *In re Bevill, Bresler & Schulman Asset Mgmt. Corp.*, 805 F.2d 120 (3d Cir. 1986), the court established a stringent test for when corporate officers can assert a personal privilege over communications with corporate counsel. The court held that the privilege belongs to the corporation, not the individual officers, and that officers cannot prevent the corporation from waiving the privilege.

The *Bevill* doctrine establishes the **Third Circuit Absolute Control Principle** for AI: the privilege belongs entirely to the attorney and the client, not to the AI platform. The AI has no independent standing to waive the privilege, and its internal processing mechanisms cannot effectuate a waiver unless the attorney explicitly authorizes it. This principle neutralizes the argument that the mere act of transmitting data to an AI platform constitutes a waiver, provided the platform is operating under an enterprise DPA that contractually subordinates the platform's data rights to the attorney's control.

G. The Supreme Court: Hickman v. Taylor and Opinion Work Product

Finally, even if a court were to find that the attorney-client privilege does not apply to AI outputs, the work product doctrine provides a near-absolute fallback defense. Work product is even more protective than privilege.

In *Hickman v. Taylor*, 329 U.S. 495 (1947), the Supreme Court established that the work product of an attorney—including interviews, statements, memoranda, correspondence, and mental impressions—is protected from discovery.

AI prompts are work product if:

- They reflect attorney mental impressions,
- They are created in anticipation of litigation,
- Or they are part of legal strategy.

Confidentiality protections strengthen the argument. If the AI platform does not retain data, does not train on prompts, and is contractually bound to confidentiality, then the prompts are opinion work product, which is nearly absolute. As applied in *Tremblay v. OpenAI*, 2024 WL 3748003 (N.D. Cal. 2024), the attorney's prompt design is protected

opinion work product. The **Hickman-Tremblay Doctrine** establishes that when an attorney structures a complex prompt to direct an AI's analysis, the attorney's mental impressions and legal theories are embedded in that prompt. The AI's output is merely the mechanical execution of the attorney's protected strategy. Therefore, both the prompt and the resulting output are protected from discovery.

Summary of the Circuit-Level Defense Matrix

Doctrine	Authority	Circuit(s)	Key Holding	Application to Agentic AI
Rejection of Artificial Limits	*Upjohn Co. v. United States*, 449 U.S. 383 (1981)	Supreme Court	Privilege must facilitate free flow of information; rigid tests rejected	AI is a necessary modern conduit for information flow
Tool Not Person	*Warner v. Gilbarco*, 2026 WL 373043 (E.D. Mich. 2026); *Columbia/HCA*, 293 F.3d 289 (6th Cir. 2002)	6th Circuit	AI programs are "tools, not persons"; waiver requires adversary disclosure	*Ackert* governs persons; AI is a tool, defeating *Ackert*
Necessary Conduit	*Harper & Row*, 423 F.2d 487 (7th Cir. 1970); *In re Grand Jury (Osman)*, 220 F.3d 568 (7th Cir. 2000)	7th Circuit	Privilege protects third-party agents translating data for legal advice	AI functions as a *Kovel* translator of complex data
Functional Equivalent	*In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994)	8th Circuit	Five-factor test extends privilege to independent consultants	Enterprise AI satisfies all five *Bieter* factors
Investigative Agent	*Admiral Ins.*, 881 F.2d 1486 (9th Cir. 1989); *Graf*, 610 F.3d 1148 (9th Cir. 2010)	9th Circuit	Protects the process of gathering information by agents	AI document analysis is protected investigative work
Absolute Control	*In re Bevill*, 805 F.2d 120 (3d Cir. 1986)	3rd Circuit	Privilege belongs to the entity/client, not the agent	AI has no personal privilege to waive; attorney controls
Opinion Work Product	*Hickman v. Taylor*, 329 U.S. 495 (1947);	Supreme Court / All	Attorney's mental impressions	Near-absolute protection for

Doctrine	Authority	Circuit(s)	Key Holding	Application to Agentic AI
	Tremblay v. OpenAI, 2024 WL 3748003 (N.D. Cal. 2024)	Circuits	embedded in prompts are opinion work product	attorney-directed AI analysis

H. The Practical Deployment Protocol

The doctrinal framework above is powerful, but it is only effective if the underlying facts support it. The following conditions must be present for the full defense matrix to apply:

1. **The AI must be operated at counsel's direction.** This is the single most important factual predicate. *Heppner* failed because the client acted alone. *Warner* succeeded because the plaintiff was acting as her own counsel. In a represented client context, all AI use for legal purposes must be directed by the attorney, not initiated by the client independently.

2. **The platform must be enterprise-tier with zero data retention.** The *Bieter* five-factor test requires that the communication "not be disseminated beyond those who need to know." A consumer AI platform with a governmental disclosure provision fails this test. An enterprise platform with a Data Processing Agreement and ZDR satisfies it.

3. **The purpose must be to facilitate legal advice, not independent analysis.** The *Kovel* doctrine requires that the third party's role be to assist the attorney in rendering legal advice, not to provide independent expertise. Attorneys should structure AI prompts to make explicit that the AI is organizing, translating, and synthesizing information for counsel's review—not generating independent legal conclusions.

4. **The AgFi framework must be applied to platform selection.** A platform that cannot demonstrate operational compliance with its contractual commitments—through SOC 2 Type II certification, ZDR architecture, and documented data handling practices—cannot satisfy the *Bieter* confidentiality requirement regardless of what its contract says. The AgFi Scorecard is the operational tool for making this determination.

H. The Ackert Counter-Argument: When AI Crosses the Line from Conduit to Independent Expert

While the *Kovel* and *Bieter* doctrines provide the shield, attorneys must be acutely aware of the sword: the *Ackert* limitation. In *United States v. Ackert*, 169 F.3d 136 (2d Cir. 1999), the Second Circuit held that the attorney-client privilege does not protect communications with a third-party consultant if the consultant is providing independent expertise or advice, rather than merely translating or facilitating the client's communication to the attorney. [3]

Opposing counsel seeking to pierce the HPS Doctrine will inevitably rely on *Ackert*. The argument will be framed as follows: *Generative AI is not merely a translator or a conduit; it is an independent analytical engine. When an attorney asks an AI to synthesize case law or draft a legal strategy, the AI is generating novel, independent expertise. Therefore, under Ackert, the communication is not privileged.*

This argument, while superficially appealing, fails upon careful analysis. To defeat the *Ackert* challenge, attorneys must structurally prevent the AI from acting as an independent expert. This is achieved through the TOLFPC Quality Assurance Protocol (detailed in Section VII). If the attorney uses the AI to generate *new* legal theories without human direction, the *Ackert* challenge may succeed. However, if the attorney uses the AI strictly to process, organize, and synthesize data *provided by the attorney or the client*, the AI remains a necessary conduit. Courts must recognize that an AI platform, lacking sentience and independent agency, cannot possess "independent expertise" in the way a human investment banker did in *Ackert*. The AI's output is entirely dependent on the attorney's prompt. Therefore, as long as the prompt reflects the attorney's mental impressions and directs the AI's processing parameters, the AI remains a tool of the attorney, not an independent expert, and the *Ackert* limitation does not apply.

The practical implication of this analysis is critical: the attorney's prompt is not merely a convenience; it is the legal instrument that determines whether the AI functions as a protected *Kovel* agent or an unprotected *Ackert* consultant. An attorney who asks an AI "What are the strongest defenses in a securities fraud case?" has asked for independent expertise. An attorney who asks an AI "Organize the following 500 documents by date and extract every reference to the defendant's knowledge of the alleged fraud" has directed a conduit. The former risks *Ackert* exposure; the latter is protected under *Bieter*. This distinction must be embedded in every AI governance policy at every law firm.

XVI. THE DAUBERT EVIDENTIARY FRAMEWORK FOR AI OUTPUTS

Beyond the question of privilege lies the inevitable evidentiary challenge: if an attorney successfully protects their AI prompts and outputs from discovery, what happens when they attempt to introduce AI-generated analysis or data synthesis as evidence in court? The HPS Doctrine protects the *process*, but the Federal Rules of Evidence govern the *product*.

As agentic AI platforms become capable of processing millions of documents to identify patterns of fraud, calculate damages, or reconstruct timelines, attorneys will increasingly seek to admit these AI-generated syntheses as evidence. This will trigger immediate challenges under Federal Rule of Evidence 702 and the *Daubert* standard.

A. The Daubert Vulnerability of Black-Box AI

Under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), expert testimony based on scientific or technical methodology must be reliable, peer-reviewed, and have a known error rate. Generative AI presents a unique *Daubert* crisis: it is inherently non-deterministic. The same prompt can produce different outputs on different days. Furthermore, the internal weighting mechanisms of commercial LLMs are proprietary "black boxes," making it impossible for opposing counsel to cross-examine the algorithm's methodology. If an attorney attempts to introduce an AI-generated financial analysis without a human expert to validate the underlying methodology, the evidence will be excluded under *Daubert* for lacking a verifiable error rate and testable methodology. The hallucination rates documented by Dahl et al. (2024) and Magesh et al. (2025) — ranging from 17.7% to 88% depending on the model and task — provide opposing counsel with precisely the kind of published error-rate data that supports a *Daubert* exclusion motion. [11] [12]

B. The Human-in-the-Loop Evidentiary Bridge

To survive a *Daubert* challenge, the AI cannot be the expert; the AI must be the *tool* used by the expert. The evidentiary bridge requires a human expert to testify that: (1) they selected the specific AI platform based on its established reliability for the specific task (e.g., using Harvey AI for legal document extraction rather than a consumer LLM);

(2) they designed the specific prompts and parameters used to process the data; (3) they independently verified a statistically significant sample of the AI's output against the primary source documents; and (4) they take professional responsibility for the final conclusions.

Under this framework, the AI is treated analogously to complex forensic software or e-discovery predictive coding algorithms (Technology-Assisted Review, or TAR). Courts have routinely accepted TAR under Fed. R. Civ. P. 26 when the methodology is transparent and validated by human review. The same standard must apply to agentic AI. The TOLFPC Protocol's requirement for independent verification (Section VII) is not just an ethical mandate; it is the foundational requirement for evidentiary admissibility. An attorney who follows the TOLFPC Protocol has simultaneously satisfied both the ethical competence standard under ABA Model Rule 1.1 and the evidentiary reliability standard under *Daubert*. This is the doctrinal bridge between legal ethics and the law of evidence that the existing literature has entirely failed to construct.

XVII. THE PRACTICAL ROADMAP FOR SAFE AI ADOPTION

The integration of agentic AI into legal practice is no longer optional; it is a competitive necessity and, increasingly, a component of the duty of competence. However, as *Heppner* demonstrates, the risks of improper adoption are severe. The following roadmap provides a step-by-step guide for law firms and corporate legal departments to safely integrate AI while preserving privilege and complying with ethical obligations.

Step 1: Establish the Governance Framework

Before any attorney or staff member uses an AI tool for client work, the firm must establish a formal AI governance policy. This policy must:

- Prohibit the use of consumer-tier AI platforms (e.g., free ChatGPT, free Claude) for any client-related work.
- Mandate the use of approved, enterprise-tier platforms with verified Zero Data Retention (ZDR) agreements.
- Require that all AI use be directed by an attorney and documented in the client file.

- Establish a mandatory training program on AI hallucination risks and the firm's verification protocols.

Step 2: Update Client Engagement Letters

As discussed in Section IV, client engagement letters must be updated immediately to address AI use. The letter must explicitly state that the client is prohibited from using consumer AI platforms to analyze their legal matter or process communications with counsel, and that doing so may result in a waiver of the attorney-client privilege. The letter should also disclose the firm's use of enterprise AI tools and obtain the client's informed consent, consistent with ABA Formal Opinion 512.

Step 3: Implement the AgFi Evaluation Protocol

The firm's IT or knowledge management department must evaluate all proposed AI platforms using the AgFi Scorecard (Section V). Only platforms scoring in the "High AgFi" range (15-18 points) should be approved for complex legal analysis. The evaluation must include a review of the platform's Data Processing Agreement, SOC 2 Type II certification, and hallucination rates on legal benchmarks.

Step 4: Deploy the TOLFPC Quality Assurance Protocol

The firm must mandate the use of the TOLFPC Quality Assurance Protocol (Section VII) for all AI-assisted drafting. This includes:

- **Phase 1:** Using the Sanguage Technique, Metatag Protocol, and Constitutional Persona during drafting.
- **Phase 2:** Executing the Post-Drafting Integrity Cascade, including manual Chain-of-Verification (CoV) review and multi-platform verification of all citations and factual propositions.

Step 5: Document the Privilege Architecture

To preserve the *Kovel*/*Bieter* and *Tremblay* defenses, attorneys must document their AI use. This documentation should include:

- The specific legal purpose for which the AI was deployed.

- The prompts used to direct the AI (which constitute protected opinion work product).
- The verification steps taken to ensure the accuracy of the AI's output.
- A record of the enterprise DPA governing the platform used.

Step 6: Monitor the Regulatory Landscape

The firm must designate an individual or committee to monitor the rapidly evolving AI regulatory landscape (Section XI), including state AI laws, FTC enforcement actions, and updates to the ABA Model Rules. The firm's AI governance policy must be updated regularly to reflect these changes.

XVIII. SUPPLEMENTAL DELIVERABLES: OPERATIONALIZING THE HPS DOCTRINE

To transition the HPS Doctrine from theory to practice, the following four operational tools are provided for immediate deployment by law firms and corporate legal departments.

A. Model Privilege Memorandum for Internal Firm Use

MEMORANDUM TO: All Attorneys and Legal Staff **FROM:** Office of the General Counsel / Risk Management Committee **DATE:** March 15, 2026 **SUBJECT:** Preservation of Attorney-Client Privilege and Work Product When Using Artificial Intelligence

1. Purpose and Scope This memorandum establishes the firm's legal position and mandatory protocols regarding the use of Artificial Intelligence (AI) platforms to ensure the preservation of attorney-client privilege and the work-product doctrine.

2. Legal Standard: The Bieter and Kovel Doctrines Under the *Bieter* doctrine (*In re Bieter Co.*, 8th Cir. 1994) and the *Kovel* doctrine (*United States v. Kovel*, 2d Cir. 1961), disclosure of privileged information to a third party does not waive privilege if the third party is acting as a "functional equivalent" of the attorney or is necessary to facilitate legal representation. Furthermore, under *Warner v. Gilbarco* (E.D. Mich. 2026), AI platforms are legally classified as "tools, not persons."

Therefore, prompts entered into a privacy-protected AI platform are protected by attorney-client privilege and the work-product doctrine, provided the platform does not retain data, does not train on prompts, and is contractually bound to confidentiality.

3. Mandatory Protocols To ensure all AI use falls within this protected framework, all personnel must adhere to the following:

- **Approved Platforms Only:** Personnel may only use enterprise-tier AI platforms explicitly approved by the firm (e.g., Harvey AI, Lexis Protégé, Claude Enterprise). The use of consumer-tier platforms (e.g., free ChatGPT, free Claude) for client work is strictly prohibited and constitutes a terminable offense.
- **Attorney Direction:** All AI use for client matters must be directed by an attorney for the specific purpose of rendering legal advice or preparing for litigation.
- **Prompt Documentation:** All complex prompts used to direct legal analysis must be saved to the client file. Under *Tremblay v. OpenAI* (N.D. Cal. 2024), these prompts constitute protected opinion work product.
- **Verification:** All AI-generated output must be independently verified using the firm's Chain-of-Verification (CoV) protocol before being incorporated into any final work product.

B. Risk-Scoring Matrix for Evaluating AI Platforms

This matrix operationalizes the AgFi framework for procurement decisions.

Risk Factor	High Risk (0 pts)	Moderate Risk (1 pt)	Low Risk (2 pts)
Data Retention	Indefinite retention; user must manually delete	30-day retention for abuse monitoring	Zero Data Retention (ZDR) architecture
Model Training	Opt-out required; default is training	Opt-in required; default is no training	Contractual prohibition on training
Third-Party Disclosure	Permits disclosure to "governmental authorities"	Requires subpoena or court order	Requires subpoena + user notification
Security Certification	None or self-attested	SOC 2 Type I	SOC 2 Type II + ISO 27001
Hallucination Rate	>30% on legal benchmarks	15% - 30% on legal benchmarks	<15% on legal benchmarks
Context Stability	Fails <20k tokens	Fails 20k-50k tokens	Stable >50k tokens

Scoring:

- **10-12 Points:** Approved for privileged legal work.
- **6-9 Points:** Approved for non-privileged administrative tasks only.
- **0-5 Points:** Prohibited from firm networks.

C. Recommended AI-Use Policy for Attorneys

1. Prohibition on Consumer AI: Attorneys shall not input any client confidential information, personally identifiable information (PII), or protected health information (PHI) into any consumer-tier AI platform. **2. Mandatory Verification:** Attorneys retain ultimate responsibility for all work product under ABA Model Rule 1.1. Attorneys must independently verify all citations, factual propositions, and legal conclusions generated by an AI tool using primary sources. **3. Client Disclosure:** Attorneys must inform clients if AI tools will be used materially in their representation, consistent with ABA Formal Opinion 512, and must include the firm's standard AI disclosure language in all new engagement letters. **4. Prompt Engineering as Work Product:** Attorneys should draft prompts with the understanding that the prompt itself is opinion work

product. Prompts should be structured to direct the AI's analysis, embedding the attorney's mental impressions and legal theories. **5. Prohibition on "Black Box" Filings:** Attorneys shall not file any document with a court that was generated entirely by an AI without substantive human review and modification.

D. Litigation-Ready Argument Outline for Asserting Privilege Over AI Prompts

If opposing counsel moves to compel production of an attorney's AI prompts or the resulting outputs, the following argument structure should be deployed:

I. The AI Platform is a Tool, Not a Person (The Sixth Circuit Defense)

- *Authority: Warner v. Gilbarco, Inc.*, 2026 WL 373043 (E.D. Mich. Feb. 10, 2026); *In re Columbia/HCA*, 293 F.3d 289 (6th Cir. 2002).
- *Argument:* Waiver requires disclosure to a third-party adversary. An enterprise AI platform is a software tool, analogous to Westlaw or a cloud-storage provider. Transmitting data to a tool does not constitute a third-party disclosure.

II. The AI Functions as a Necessary Conduit (The Seventh/Eighth Circuit Defense)

- *Authority: In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994); *In re Grand Jury Proceedings (Osman)*, 220 F.3d 568 (7th Cir. 2000).
- *Argument:* Even if the AI is viewed as a third-party agent, it satisfies the *Bieter* five-factor test. It is deployed at counsel's direction, processes relevant information, and operates under a strict confidentiality agreement (ZDR) that prevents dissemination. It is the functional equivalent of a protected consultant.

III. The Prompts Constitute Opinion Work Product (The Supreme Court Defense)

- *Authority: Hickman v. Taylor*, 329 U.S. 495 (1947); *Tremblay v. OpenAI, Inc.*, 2024 WL 3748003 (N.D. Cal. Aug. 8, 2024).
- *Argument:* The prompts drafted by counsel contain counsel's mental impressions, legal theories, and strategic directives. Under *Tremblay*, these prompts are near-absolutely protected opinion work product, regardless of the AI's output.

IV. The Heppner Decision is Factually Inapposite

- *Authority: United States v. Heppner*, No. 25 Cr. 503 (JSR) (S.D.N.Y. Feb. 17, 2026).

XII.A. THE HEPPNER PROTECTIVE SHIELDS: DEMONSTRATIVE ANTI-CONTEXT DEGRADATION ALG

A. Purpose and Theoretical Basis

The Heppner Protective Shields are designed to mitigate context degradation, anchoring bias, and multi-turn reasoning drift in AI-assisted legal analysis. Empirical research demonstrates that large language models exhibit measurable performance decline in extended conversational chains, particularly where early assumptions propagate unchallenged. In a comprehensive evaluation of multi-turn conversations involving over 200,000 simulated exchanges, researchers found that LLMs experience an average performance drop of 39% across generation tasks when moving from single-turn to multi-turn settings, largely because "when LLMs take a wrong turn in a conversation, they get lost and do not recover." Philippe Laban, Hiroaki Hayashi, Yingbo Zhou & Jennifer Neville, LLMs Get Lost In Multi-Turn Conversation, arXiv:2505.06120 (May 9, 2025). [67]

This degradation is compounded by the "Lost in the Middle" phenomenon, where models fail to robustly utilize relevant information located in the middle of long input contexts, with performance degrading significantly even for models explicitly designed for long-context tasks. Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni & Percy Liang, Lost in the Middle: How Language Models Use Long Contexts, 12 Transactions of the Ass'n for Computational Linguistics 157 (2024). [68] Furthermore, LLMs exhibit pronounced anchoring bias, where initial information disproportionately shapes subsequent outputs, making self-correction difficult across all tested models including GPT-4, Claude 2, Gemini Pro, and GPT-3.5. Jeremy K. Nguyen, Human Bias in AI Models? Anchoring Effects and Mitigation Strategies in Large Language Models, 43 J. Behavioral & Experimental Finance 100971 (2024). [69]

To counteract this phenomenon, the Protective Shields impose a structured fragmentation-and-reintegration protocol, ensuring that each analytical output is independently validated before incorporation into a unified litigation strategy. This approach aligns with recent findings that separating production and review into distinct, isolated sessions -- a method termed Cross-Context Review (CCR) -- significantly improves LLM output quality by breaking the anchoring effect of the original conversation history, achieving an F1 score of 28.6% versus 24.6% for same-session self-review (p=0.008, d=0.52). Tae-Eun Song, Cross-Context Review: Improving LLM Output Quality by Separating Production and Review Sessions, arXiv:2603.12123 (Mar. 12, 2026). [70]

B. Core Principle: Segmented Cognition with Adversarial Reintegration

The system operates on three principles: (1) Context Isolation -- prevent inherited analytical bias by strictly limiting the factual and legal context provided to the AI in any single session; (2) Adversarial Duplication -- force independent challenge of conclusions by requiring a separate, isolated AI session to attack the initial output; and (3) Controlled Reintegration -- merge only validated outputs into the master theory, ensuring that reasoning drift is caught before it contaminates the broader litigation strategy.

C. Demonstrative Algorithm (Heppner Shield Protocol v1.0)

The following pseudo-code illustrates the operational workflow of the Heppner Shield Protocol:

```
INPUT:
    CaseFacts F; ProceduralPosture P; LegalIssues I = {i1, i2, ..., in}
```

```
INITIALIZE:
    MasterRecord M = empty; ValidatedOutputs V = empty

STEP 1: Context Minimization
    For each issue i in I:
        Construct MinimalContext Ci = Extract(F, P, i) [essential facts only]

STEP 2: Independent Analysis (Isolation Phase)
    For each Ci: Open NewSession Si
        Prompt: "Analyze issue i under facts Ci.
                 Identify strongest arguments, weaknesses, controlling law."
        Output Ai = Result(Si)

STEP 3: Adversarial Challenge Phase
    For each Ai: Open NewSession Si-prime
        Prompt: "Assume Ai is incorrect or incomplete.
                 Identify all errors, omissions, counterarguments."
        Output Bi = Result(Si-prime)

STEP 4: Cross-Validation
    For each pair (Ai, Bi):
        If Ai withstands Bi: Mark Vi = Ai as Validated
        Else: Refine Ai to Ai*; Repeat Steps 3-4

STEP 5: Reintegration
    For each Validated Vi: Insert into MasterRecord M with:
        - Supporting authority
        - Identified vulnerabilities
        - Strategic use (motion, opposition, settlement leverage)

STEP 6: Global Consistency Check
    Open NewSession S_final
    Prompt: "Review MasterRecord M for inconsistencies,
             legal inaccuracies, and strategic weaknesses."
    Output C = ConsistencyReport
    If C identifies defects: Iterate refinement loop

OUTPUT: FinalStrategy M (fully validated, adversarially tested)
```

D. Functional Effect

The Heppner Shield Protocol produces three measurable improvements in AI-assisted legal work product: (1) higher analytical integrity through reduced anchoring to early errors; (2) increased adversarial robustness through pre-litigation stress testing; and (3) improved judicial persuasiveness because arguments that have been adversarially tested are more likely to withstand judicial scrutiny. Critically, the algorithm transforms AI from a linear drafting assistant into a multi-agent analytical system, approximating the internal review dynamics of a high-functioning litigation team.

E. Application in Complex Litigation

In complex litigation -- such as RICO enterprise allegations, arbitration challenges, or multi-claim housing proceedings -- the Protective Shields serve four functions: (1) prevent contamination of legal theories across claims by isolating each claim's analysis in a separate session; (2) isolate statutory interpretation from factual narrative bias; (3) enable parallel testing of mutually inconsistent arguments, preserving strategic optionality until final integration; and (4) allow counsel to select the strongest validated theory for each claim before committing to a final litigation strategy.

F. Conclusion

The Heppner Protective Shields represent a procedural safeguard against a newly recognized category of analytical risk: AI-induced reasoning drift in extended workflows. By enforcing

isolation, adversarial testing, and structured reintegration, the protocol ensures that legal outputs remain precise, resilient, and litigation-ready. The empirical foundation for this protocol -- grounded in peer-reviewed research on LLM context degradation [67], anchoring bias [69], and cross-context review [70] -- provides the scientific basis for treating the Heppner Shield Protocol as a proposed best practice for AI-assisted legal work under ABA Model Rule 1.1.

- *Argument: Heppner* involved a client acting alone on a consumer platform with a privacy policy permitting governmental disclosure. Here, counsel directed the use of an enterprise platform with a Zero Data Retention agreement. The confidentiality failures present in *Heppner* do not exist in this case.

XIX. CONCLUSION

The *Heppner* decision is not a prohibition on the use of AI in legal practice; it is a warning about the consequences of using the wrong tools in the wrong way. By understanding the specific factual failures that drove the ruling—the lack of attorney direction, the use of a consumer platform with a governmental disclosure provision, and the absence of a secure confidentiality architecture—attorneys can design workflows that avoid these pitfalls.

The Heppner Protective Shield (HPS) Doctrine provides a clear, operational framework for preserving attorney-client privilege and work product protections in the era of agentic AI. By combining the *Kovel* and *Bieter* doctrines' requirement for attorney direction with the technological necessity of enterprise-tier Zero Data Retention agreements, the HPS Doctrine ensures that AI functions as a protected extension of the attorney's analytical capacity, rather than an independent third party that shatters confidentiality.

Furthermore, the integration of the Agentic Fidelity (AgFi) framework and the TOLFPC Quality Assurance Protocol ensures that the use of AI meets the highest standards of professional competence, mitigating the risk of hallucination and the severe sanctions that follow.

The legal profession is at an inflection point. The forced convergence of traditional legal publishing and Silicon Valley AI development is creating tools of unprecedented power. Attorneys who master these tools within a rigorous, defensible privilege architecture will define the future of the practice. Those who ignore the architectural requirements of privilege in the pursuit of convenience will find themselves, like Bradley Heppner, stripped of their most fundamental protections.

XX. JUDICIAL CHECKLIST: RULING ON MOTIONS TO COMPEL AI COMMUNICATIONS

The following checklist is designed as a practical tool for federal judges and magistrate judges confronted with a motion to compel production of AI-generated communications in civil or criminal proceedings. It operationalizes the HPS Doctrine into a sequential, binary decision framework that can be applied at the hearing stage without requiring expert testimony. Each question maps directly to the controlling case law identified in this article.

FEDERAL JUDICIAL CHECKLIST FOR AI PRIVILEGE DETERMINATIONS *Heppner Protective Shield (HPS) Doctrine — March 2026*

STEP 1 — Platform Architecture Inquiry (The Confidentiality Gate)

> *Question: Did the attorney or client use a consumer-tier AI platform (i.e., a platform whose terms of service permit retention of user inputs, use of inputs for model training, or disclosure of inputs to governmental regulatory authorities)?*

- **YES** → Privilege is destroyed. No further analysis required. *Heppner* applies directly. The governmental disclosure provision eliminates any reasonable expectation of confidentiality. *GRANT* the motion to compel.
- **NO (Enterprise ZDR tier confirmed)** → Proceed to Step 2.

Controlling Authority: United States v. Heppner, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026); *In re Bieter Co.*, 16 F.3d 929, 935 (8th Cir. 1994) (communication must "not be disseminated beyond those who need to know").

STEP 2 — Attorney Direction Inquiry (The Agency Gate)

> *Question: Was the AI used at the direction of, or under the supervision of, licensed counsel for the purpose of facilitating legal advice or preparing for litigation?*

- **NO (Client acted independently, as in *Heppner*)** → Privilege is destroyed. The *Kovel* agency relationship is absent. *GRANT* the motion to compel.
- **YES (Attorney directed or supervised AI use)** → Proceed to Step 3.

Controlling Authority: United States v. Kovel, 296 F.2d 918, 922 (2d Cir. 1961); *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026) (privilege destroyed where client acted "on his own volition").

STEP 3 — The Ackert Fracture Line Inquiry (The Conduit vs. Expert Gate)

> *Question: Was the AI used as a conduit to process, organize, translate, or synthesize information provided by the attorney or client — or was it used to generate independent legal analysis, strategy, or expertise?*

- **Independent Expert Mode (AI generated novel legal theories or strategy without attorney direction)** → Privilege may be defeated under *Ackert*. Conduct *in camera* review of the specific prompts to determine the degree of attorney direction. If prompts reflect attorney mental impressions, proceed to Step 4. If prompts are generic requests for independent analysis, *GRANT* the motion to compel as to those specific documents.
- **Conduit Mode (AI organized, translated, or synthesized attorney-directed data)** → Proceed to Step 4.

Controlling Authority: United States v. Ackert, 169 F.3d 136, 139 (2d Cir. 1999) (privilege does not extend to third party providing "independent expertise"); *In re Bieter Co.*, 16 F.3d at 937 (privilege extends to consultant acting as "functional equivalent" of employee).

STEP 4 — Work Product Doctrine Inquiry (The Mental Impressions Gate)

> *Question: Do the AI-generated documents reflect the attorney's mental impressions, conclusions, opinions, or legal theories?*

- **NO (Documents are purely factual summaries without attorney direction)** → Ordinary work product protection applies. Subject to substantial need showing by opposing party under Fed. R. Civ. P. 26(b)(3)(A). Conduct standard substantial need analysis.
- **YES (Documents reflect attorney mental impressions embedded in prompts or outputs)** → Opinion work product protection applies. Protection is near-absolute under *Hickman v. Taylor*, 329 U.S. 495 (1947), and *Upjohn Co. v. United States*, 449 U.S. 383 (1981). *DENY* the motion to compel as to these documents.

Controlling Authority: Hickman v. Taylor, 329 U.S. 495, 510–11 (1947); *Upjohn Co. v. United States*, 449 U.S. 383, 400 (1981); *Tremblay v. OpenAI, Inc.*, 2024 WL 3748003, at *4 (N.D. Cal. Aug. 8, 2024) (AI prompts reflecting attorney strategy are opinion work product).

STEP 5 — The Bieter Functional Equivalent Confirmation

> *Question: Does the enterprise AI platform satisfy the Bieter five-factor test: (1) purpose was to secure legal advice; (2) AI processed information relevant to the legal matter; (3) use was directed by counsel; (4) processing was the AI's specific function; and (5) information was not disseminated beyond those who need to know?*

- **ALL FIVE FACTORS SATISFIED** → The AI platform is the functional equivalent of a protected agent. Privilege is preserved. *DENY* the motion to compel.
- **ONE OR MORE FACTORS NOT SATISFIED** → Conduct *in camera* review of the specific documents to determine which fail the *Bieter* test. Order production only of documents failing the test.

Controlling Authority: In re Bieter Co., 16 F.3d 929, 937–38 (8th Cir. 1994); *United States v. Graf*, 610 F.3d 1148, 1159 (9th Cir. 2010) (adopting *Bieter* test).

SUMMARY DISPOSITION TABLE

Scenario	Applicable Rule	Disposition
Consumer AI, no ZDR, client acted alone	*Heppner*	Grant motion to compel
Enterprise ZDR, client acted alone	*Heppner* / *Kovel*	Grant motion to compel
Consumer AI, attorney directed	*Heppner* (confidentiality failure)	Grant motion to compel
Enterprise ZDR, attorney directed, conduit use	*Kovel* / *Bieter*	Deny motion to compel
Enterprise ZDR, attorney directed, independent expert use	*Ackert*	In camera review
Enterprise ZDR, attorney directed, mental impressions in prompts	*Hickman* / *Tremblay*	Deny motion to compel (opinion work product)

XXI. ATTORNEY CERTIFICATION OF AI COMPLIANCE WITH THE HPS DOCTRINE

The following template is designed for immediate use by attorneys who have used agentic AI in the preparation of any brief, motion, memorandum, or other court filing. It may be attached as an exhibit to any filing or submitted as a standalone certification in response to a court order or opposing counsel's discovery demand. It operationalizes the HPS Doctrine into a sworn, litigation-ready document.

CERTIFICATION OF COMPLIANCE WITH THE HEPPNER PROTECTIVE SHIELD (HPS) DOCTRINE Regarding Use of Artificial Intelligence in Preparation of [DOCUMENT TITLE]

I, [ATTORNEY NAME], [BAR NUMBER], counsel for [PARTY NAME] in the above-captioned matter, hereby certify as follows:

1. Platform Architecture Compliance. All artificial intelligence platforms used in the preparation of the above-referenced document were accessed exclusively through enterprise-tier application programming interfaces (APIs) or enterprise subscription agreements that include: (a) Zero Data Retention (ZDR) provisions prohibiting the

platform from retaining user inputs or AI outputs beyond the duration of the active session; (b) Data Processing Agreements (DPAs) or equivalent contractual instruments prohibiting the use of user inputs for model training; and © express prohibitions on the disclosure of user inputs to governmental regulatory authorities or any third parties without a valid legal process. No consumer-tier AI platform was used in the preparation of this document.

2. Attorney Direction and Supervision. All use of artificial intelligence in the preparation of this document was initiated, directed, and supervised by undersigned counsel. No AI platform was used by the client or any non-attorney independently to generate content that was incorporated into this document without counsel's review and direction. All AI use was for the purpose of facilitating the rendering of legal advice and the preparation of this filing, consistent with the *Kovel* doctrine (*United States v. Kovel*, 296 F.2d 918 (2d Cir. 1961)) and the functional equivalent doctrine (*In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994)).

3. Conduit Use — No Independent Expert Mode. All AI platforms were used as necessary conduits to process, organize, synthesize, and translate information provided by counsel or the client. No AI platform was used to generate independent legal theories, strategies, or conclusions without counsel's direction and review. All AI-generated content was reviewed and verified by undersigned counsel prior to inclusion in this document, consistent with the requirements of ABA Model Rule 1.1 (Competence).

4. Hallucination Verification — TOLFPC Protocol Compliance. All legal citations, case holdings, statutory provisions, and factual assertions generated or identified with the assistance of artificial intelligence have been independently verified by undersigned counsel against primary source documents, including Westlaw, LexisNexis, official government publications, or original court opinions. No AI-generated citation has been included in this document without independent verification. This certification satisfies the proposed minimum standard of assisted legal work established by the TOLFPC Quality Assurance Protocol and ABA Formal Opinion 512 (July 29, 2024).

5. Privilege Documentation. Undersigned counsel maintains a privilege log documenting: (a) the specific AI platform(s) used; (b) the enterprise tier and ZDR status of each platform; © the date and nature of each AI interaction; (d) the attorney who directed each AI interaction; and (e) the verification steps taken for each AI-generated output. This log is available for *in camera* review upon court order.

6. Accuracy Certification. Undersigned counsel certifies that the contents of this filing are accurate to the best of counsel's knowledge, information, and belief, formed after reasonable inquiry, and that the use of artificial intelligence in its preparation does not diminish counsel's personal professional responsibility for the accuracy and completeness of this document under Fed. R. Civ. P. 11 and the applicable rules of professional conduct.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on: _____

Signature: _____

Printed Name: _____

Bar Number: _____

Firm: _____

Address: _____

Phone: _____

Email: _____

XXII. GLOSSARY OF KEY TERMS

The following glossary provides formal definitions of every technical, legal, and operational term used in this manuscript. It is designed to serve as a reference for courts, practitioners, and scholars who may be unfamiliar with the specialized vocabulary at the intersection of artificial intelligence and privilege law.

Agentic AI (Agentic Artificial Intelligence) A class of artificial intelligence systems capable of autonomously planning, executing, and iterating multi-step tasks across multiple tools, data sources, and software environments without continuous human intervention at each step. Unlike conventional AI chatbots that respond to single prompts, agentic AI systems can browse the internet, write and execute code, manage files, send communications, and complete complex workflows end-to-end. The world's first commercially available agentic AI platform was Manus AI, launched in March 2025.

The legal significance of agentic AI is that its autonomous, multi-step operation creates a far larger surface area of potential privilege exposure than single-turn chatbot interactions.

Agentic Fidelity (AgFi) A proprietary framework developed by Roland G. Ottley, Esq., PA-C, of The Ottley Law Firm, P.C., for evaluating the reliability, accuracy, and privilege-compatibility of agentic AI platforms for use in legal practice. The AgFi framework assesses platforms across six dimensions: (1) Data Retention Architecture; (2) Model Training Opt-Out; (3) Governmental Disclosure Provisions; (4) SOC 2 Type II Certification; (5) Hallucination Rate; and (6) Context Stability. Platforms are scored on a 12-point scale, with scores of 10 or above qualifying for use in privileged legal work. The AgFi framework is the operational tool for satisfying the fifth element of the HPS Privilege Test.

Ackert Limitation The doctrinal boundary established by *United States v. Ackert*, 169 F.3d 136 (2d Cir. 1999), which holds that the attorney-client privilege does not extend to communications with a third-party consultant who provides independent expertise or advice, as distinguished from a consultant who merely translates or facilitates the client's communication to the attorney. In the AI context, the *Ackert* limitation is triggered when an attorney uses an AI platform to generate independent legal theories or strategies without attorney direction, rather than using the AI as a conduit to process attorney-directed data.

Attorney-Client Privilege A common law evidentiary privilege that protects confidential communications between a licensed attorney and a client made for the purpose of obtaining or rendering legal advice. The privilege belongs to the client and may be waived by the client's voluntary disclosure of the communication to a third party who is not within the scope of the privilege. In the AI context, the central question is whether disclosure of privileged information to an AI platform constitutes a waiver-triggering disclosure to a third party, or whether the AI is a protected agent of the attorney under the *Kovel* and *Bieter* functional equivalent doctrines.

Bieter Functional Equivalent Doctrine The doctrine established by *In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994), which holds that the attorney-client privilege extends to communications between counsel and an independent third-party consultant if the consultant is the "functional equivalent" of an employee or agent necessary to facilitate legal representation. The *Bieter* court established a five-factor test: (1) the purpose of the communication was to secure legal advice; (2) the consultant possessed relevant information; (3) the communication was made at the direction of

counsel; (4) the subject matter of the communication was within the consultant's specific duties; and (5) the communication was not disseminated beyond those who need to know. In the HPS Doctrine, enterprise AI platforms that satisfy all five *Bieter* factors are treated as protected agents of the attorney.

Consumer AI Tier The publicly accessible, subscription-free or low-cost version of a commercial AI platform, governed by consumer terms of service that typically permit the platform to retain user inputs, use inputs for model training, and disclose inputs to governmental regulatory authorities. The *Heppner* decision established that use of a consumer AI tier destroys attorney-client privilege because the governmental disclosure provision eliminates any reasonable expectation of confidentiality. *See United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026).

Context Window The maximum amount of text (measured in tokens, where one token approximates 0.75 words) that an AI model can process in a single interaction. As of March 2026, leading commercial models have context windows ranging from 128,000 tokens (GPT-4o) to 2 million tokens (Gemini 1.5 Pro). The legal significance of context window size is that larger context windows allow attorneys to input entire case files, deposition transcripts, or document productions for AI analysis — but also increase the risk of "lost in the middle" degradation, where the AI's attention to information in the middle of a long context is statistically weaker than its attention to information at the beginning and end. *See* Liu et al., "Lost in the Middle," arXiv:2307.03172 (2023). [9]

Data Processing Agreement (DPA) A contractual instrument between a data controller (the law firm or attorney) and a data processor (the AI platform) that governs the terms under which the processor may handle personal data on behalf of the controller. Under the EU General Data Protection Regulation (GDPR) and equivalent state privacy laws, a DPA is legally required whenever personal data is processed by a third party. In the HPS Doctrine, a DPA is a necessary but not sufficient condition for privilege protection — it must be accompanied by a Zero Data Retention provision and a prohibition on governmental disclosure to satisfy the *Bieter* confidentiality requirement.

Enterprise AI Tier The business or professional subscription version of a commercial AI platform, governed by enterprise terms of service that typically include: Zero Data Retention (ZDR) provisions; Data Processing Agreements (DPAs); prohibitions on the use of inputs for model training; and express prohibitions on disclosure of inputs to governmental authorities without legal process. The HPS Doctrine requires enterprise-tier access as a necessary condition for privilege protection.

Hallucination (AI Hallucination) The phenomenon by which a large language model generates factually incorrect, fabricated, or nonsensical information with apparent confidence. In the legal context, hallucination most commonly manifests as the generation of false case citations, fabricated judicial holdings, or invented statutory provisions. The empirical literature documents hallucination rates ranging from 17.7% (Dahl et al., 2024) to 88% (Magesh et al., 2025) depending on the model and task. Sanctions for filing AI-hallucinated citations have been imposed in *Mata v. Avianca*, *Park v. Kim*, *Wadsworth v. Walmart*, *Lacey v. State Farm*, *Johnson v. Dunn*, and *Flycatcher Corp. v. Affable Avenue*.

Heppner Protective Shield (HPS) Doctrine The unified doctrinal framework formalized in this article for preserving attorney-client privilege and work product protections when utilizing agentic AI systems. The HPS Doctrine synthesizes the *Kovel* agency doctrine, the *Bieter* functional equivalent doctrine, the *Ackert* limitation, the *Tremblay* prompt protection, and the *Hickman*/*Upjohn* work product doctrine into a five-element privilege test: (1) Attorney Direction; (2) Confidentiality Architecture (ZDR); (3) Agentic Fidelity Verification (AgFi); (4) Secure Prompt Channel; and (5) Privilege Documentation. The doctrine is named for *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026), which provided the factual template for understanding when privilege is destroyed by AI use.

Kovel Doctrine The doctrine established by *United States v. Kovel*, 296 F.2d 918 (2d Cir. 1961), which holds that the attorney-client privilege may extend to communications between a client and a non-attorney third party if the third party is acting as an agent of the attorney to facilitate the rendering of legal advice. The paradigmatic *Kovel* agent is an accountant retained by a law firm to translate complex financial data into terms the attorney can use to advise the client. In the HPS Doctrine, an enterprise AI platform directed by counsel to process, organize, and synthesize legal information is a *Kovel* agent.

Large Language Model (LLM) A class of artificial intelligence system trained on massive datasets of text to predict and generate human-like language. Commercial LLMs underlying major legal AI platforms include GPT-4o and o3 (OpenAI), Claude 3.5 Sonnet and Claude 3.7 (Anthropic), Gemini 1.5 Pro and 2.0 Flash (Google), and Llama 4 (Meta). The legal significance of LLMs is that they are inherently probabilistic, non-deterministic systems — the same prompt can produce different outputs on different occasions — which creates both the hallucination risk and the *Daubert* evidentiary challenge addressed in this article.

Opinion Work Product The category of work product that reflects the attorney's mental impressions, conclusions, opinions, or legal theories. Opinion work product receives near-absolute protection under *Hickman v. Taylor*, 329 U.S. 495 (1947), and *Upjohn Co. v. United States*, 449 U.S. 383 (1981), and cannot be compelled even upon a showing of substantial need. In the AI context, the *Tremblay* court held that AI prompts reflecting the attorney's legal strategy and mental impressions constitute opinion work product. This is the strongest available protection for AI-assisted legal work.

Prompt (AI Prompt) The input text, instructions, or data provided by a user to an AI platform to direct its output. In the legal context, an attorney's prompt to an AI platform is the functional equivalent of an instruction to a paralegal or associate. Under the *Tremblay* decision, prompts that reflect the attorney's mental impressions, legal strategy, or case theory are protected as opinion work product. Under the *Ackert* limitation, prompts that request independent legal analysis without attorney direction may not be protected.

Secure Prompt Channel The fourth element of the HPS Privilege Test. A secure prompt channel requires that all AI interactions involving privileged client information be conducted through an enterprise API or enterprise subscription that: (a) encrypts all data in transit and at rest; (b) does not log or retain prompts or outputs; © does not share data with third-party model trainers; and (d) provides the attorney with a contractual right to audit the platform's data handling practices. The secure prompt channel is the technical implementation of the *Bieter* confidentiality requirement.

TOLFPC Quality Assurance Protocol (TOLFPC Protocol) The Three-Part Anti-Hallucination Protocol developed by The Ottley Law Firm, P.C. (TOLFPC), establishing the proposed minimum standard of competence for AI-assisted legal work 1.1. The three phases are: (1) Pre-Submission Verification — independent verification of every AI-generated citation against primary sources before filing; (2) Cross-Platform Verification — confirmation of AI outputs using a second, independent AI platform or legal research database; and (3) Human Expert Review — review by a licensed attorney with subject-matter expertise before any AI-generated analysis is incorporated into a court filing. The TOLFPC Protocol simultaneously satisfies the ethical competence standard under Rule 1.1 and the evidentiary reliability standard under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).

Work Product Doctrine A qualified evidentiary protection established by *Hickman v. Taylor*, 329 U.S. 495 (1947), and codified in Fed. R. Civ. P. 26(b)(3), that protects documents and tangible things prepared in anticipation of litigation or for trial by or

for a party or its representative. Unlike the attorney-client privilege, the work product doctrine does not require an attorney-client relationship — it protects any document prepared by a party or its representative in anticipation of litigation. In the AI context, AI-generated documents may qualify for work product protection if they were prepared at the direction of counsel in anticipation of litigation and reflect counsel's mental impressions.

Zero Data Retention (ZDR) A contractual and architectural commitment by an AI platform provider that user inputs and AI outputs will not be retained, stored, or logged beyond the duration of the active session. ZDR is the single most important technical requirement for privilege protection under the HPS Doctrine. Without ZDR, the AI platform retains a copy of every privileged communication, creating a permanent third-party repository of privileged information that is subject to subpoena, government demand, and data breach. Enterprise ZDR agreements typically include: session-level data deletion; no model training on user inputs; no logging of prompts or outputs; and contractual indemnification for unauthorized disclosure.

REFERENCES

[1] *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026). [2] *United States v. Kovel*, 296 F.2d 918 (2d Cir. 1961). [3] *United States v. Ackert*, 169 F.3d 136 (2d Cir. 1999). [4] Warner v. Gilbarco, No. 2:24-cv-12333, 2026 WL 373043 (E.D.
decision). [4a] *In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994). [5] *Tremblay v. OpenAI, Inc.*, 2024 WL 3748003 (N.D. Cal. Aug. 8, 2024). [6] Roland G. Ottley, *The Consumer Guide to Quality Agentic AI Platforms: Agentic Fidelity Guidance (AgFi)*, The Ottley Law Firm, P.C. & Manus AI (March 2026). [7] "Intelligence Degradation in Large Language Models," AI Research Institute (2026). [8] Chroma Research, "Attention Distribution in Extended Context Windows" (2025). [9] Nelson F. Liu et al., "Lost in the Middle: How Language Models Use Long Contexts," arXiv:2307.03172 (2023). [10] Illinois Attorney Registration and Disciplinary Commission (ARDC), *2026 Report on AI in Legal Practice* (2026). [11] Varun Magesh et al., "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools," *Journal of Empirical Legal Studies*, Vol. 22, Issue 2, pp. 216–242 (2025). [12] Matthew Dahl et al., "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models," *Journal of Legal Analysis*, Vol. 16, Issue 1, pp. 64–93 (2024). [13] TechCrunch, "OpenAI's New Reasoning Models Show Higher Hallucination Rates

on PersonQA" (2025). [14] *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. June 22, 2023). [15] *Park v. Kim*, 91 F.4th 610 (2d Cir. Jan. 30, 2024). [16] *Wadsworth v. Walmart, Inc.*, 348 F.R.D. 489 (D. Wyo. Feb. 24, 2025). [17] *Lacey v. State Farm Gen. Ins. Co.*, 2025 WL 1363069 (C.D. Cal. May 6, 2025). [18] *Johnson v. Dunn*, No. 2:21-cv-01701 (N.D. Ala. July 23, 2025). [19] Flycatcher Corp. Ltd. v. Affable Avenue LLC, No.

(S.D.N.Y. Feb. 5, 2026). [20] American Bar Association, Formal Opinion 512: Generative Artificial Intelligence Tools (July 29, 2024). [21] American Bar Association Task Force on Law and Artificial Intelligence, *Year 2 Report on the Impact of AI on the Practice of Law* (Dec. 15, 2025). [22] Vectara Hallucination Leaderboard (2026). [23] Microsoft, "Data, Privacy, and Security for Microsoft 365 Copilot" (2026). [24] Harvey AI, "Security and Privacy Architecture" (2026). [25] TechCrunch, "Legal AI startup Harvey confirms $8B valuation" (Dec. 4, 2025). [26] Bloomberg, "Meta Acquires Manus AI for $2 Billion" (Dec. 2025). [27] *New York Times Co. v. OpenAI*, No. 1:23-cv-11195 (S.D.N.Y. Jan. 5, 2026) (Discovery Order). [28] Carolyn Elefant, "ChatGPT Deep Research vs. Lexis+ AI: A Comparative Test," MyShingle.com (May 20, 2025). [29] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Dec. 11, 2025). [30] Children and Teens' Online Privacy Protection Act (COPPA 2.0), S.836 (Passed Senate Mar. 5, 2026). [31] Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks (TAKE IT DOWN) Act, Pub. L. No.

(Signed Dec. 19, 2025). [33] California Transparency in Frontier AI Act, SB 53 (Signed Sept. 29, 2025). [34] Texas Responsible AI Governance Act (TRAIGA), HB 149 (Signed June 22, 2025). [35] European Union Artificial Intelligence Act (Entry into force Aug. 1, 2024). [36] *FTC v. DoNotPay*, No. 232-3042 (Final Order Jan. 16, 2025). [37] *FTC v. Rytr*, No. 232-3052 (Order Set Aside Dec. 22, 2025). [38] Italian Data Protection Authority (Garante), Decision regarding OpenAI (Dec. 20, 2024). [39] *Upjohn Co. v. United States*, 449 U.S. 383 (1981). [40] *Harper & Row Publishers, Inc. v. Decker*, 423 F.2d 487 (7th Cir. 1970). [41] *In re Grand Jury Proceedings (Osman)*, 220 F.3d 568 (7th Cir. 2000). [42] *In re Bieter Co.*, 16 F.3d 929 (8th Cir. 1994). [43] Harvey AI Engineering Blog, "Agentic Search in Legal Workflows" (2026). [44] *In re Bevill, Bresler & Schulman Asset Mgmt. Corp.*, 805 F.2d 120 (3d Cir. 1986). [45] VLAIR Benchmark Study on Legal AI Performance (2025). [46] Thomson Reuters, "Rebuilding for the Agent Era: The Next Generation of CoCounsel Legal" (Mar. 10, 2026). [47] LexisNexis Press Release, "LexisNexis Enhances Lexis+ with Protégé Platform by Integrating Anthropic Legal Plugin" (Feb. 24, 2026). [48] LexisNexis Press Release, "LexisNexis & Harvey Announce Strategic Alliance" (June 18, 2025). [49] Harvey AI Press Release, "Harvey Announces Integration with Microsoft 365 Copilot" (Mar. 2026). [50] arXiv, "Predatory Consent: How Free AI Applications Exploit Low-Literacy Users" (Jan. 2026). [51] *Admiral Ins. Co. v. U.S. Dist. Court for the*

Dist. of Ariz., 881 F.2d 1486 (9th Cir. 1989). [52] Pew Research Center, “Teens and AI Chatbots” (2025); EPIC Survey on AI Companions (2025). [53] *Doe v. Character.AI*, Complaint (M.D. Fla. Oct. 2024). [54] Reuters, “Google and Character.AI Settle Teen Suicide Lawsuits” (Jan. 2026). [55] Federal Trade Commission, Section 6(b) Orders to AI Companies Regarding Child Privacy (Sept. 2025). [56] Federal Trade Commission, Final Rule Amending COPPA (Jan. 2025). [57] National Association of Attorneys General, Bipartisan Letter to AI Developers on Child Safety (Aug. 2025). [58] *United States v. Graf*, 610 F.3d 1148 (9th Cir. 2010). [59] *In re Columbia/HCA Healthcare Corp. Billing Practices Litigation*, 293 F.3d 289 (6th Cir. 2002). [60] *Hickman v. Taylor*, 329 U.S. 495 (1947).

[71] Fortis Advisors, LLC v. Krafton, Inc., C.A. No. 2025-0805-LWW (Del. Ch. Mar. 16, 2026) (Will, V.C.), Post-Trial Opinion. [72] Id. at 2 (describing CEO Changhan Kim's use of ChatGPT to formulate a corporate takeover strategy to avoid a (million earnout payment). [73] Id. at 15 (quoting the court's finding that Kim 'turned to ChatGPT for help' after internal executives warned that firing the founders without cause might trigger a lawsuit). [74] Id. at 16 (describing the AI chatbot's preparation of a 'Response Strategy to a 'No-Deal' Scenario,' including a 'pressure and leverage package' and an 'implementation roadmap by scenario,' which Krafton subsequently executed).