THE HEPPNER PROTECTIVE SHIELD:

Agentic Fidelity, Zero Data Retention, and the Cognitive Protection of Prompts as Opinion Work Product in the Age of Artificial Intelligence

By Roland G. Ottley, Esq., PA-C *Principal Attorney | The Ottley Law Firm, P.C.*

> ***Author's Note on Case Citations:*** *All cases cited in this manuscript are real judicial decisions, not hypotheticals. United States v. Heppner, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026) (Rakoff, J.), and Warner v. Gilbarco, Inc., No. 2:24-cv-12333, 2026 WL 373043 (E.D. Mich. Feb. 10, 2026) (Patti, M.J.), are decisions issued in February 2026. Their recency may cause some readers to question their existence; the full citations and secondary commentary are provided in the References section. Tremblay v. OpenAI and Concord Music Group v. Anthropic are similarly real pending or decided matters. Concord Music Group, Inc. v. Anthropic PBC, originally filed in M.D. Tenn., was transferred to N.D. Cal., where the May 2025 work product ruling was issued under docket No. 24-cv-03811-EKL (SVK), 2025 WL 1482734. The manuscript was finalized in March 2026.*

> ***Doctrinal Status Legend:*** *This manuscript draws on three categories of legal authority, which are distinguished throughout as follows:*
>
> ***[SETTLED LAW]*** *— Propositions supported by binding Supreme Court or circuit court authority that is well-established and not subject to reasonable dispute. Examples: Hickman v. Taylor, Upjohn, Kovel, FRCP 26(b)(3).*
>
> ***[EMERGING AUTHORITY]*** *— Propositions supported by recent district court decisions, magistrate rulings, or a developing but not yet uniform body of authority. These propositions are persuasively supported but have not yet achieved the status of settled law. Examples: Tremblay v. OpenAI, Warner v. Gilbarco, United States v. Heppner.*

> ***[PROPOSED FRAMEWORK]*** *— Analytical constructs and normative arguments advanced by this manuscript as frameworks that courts and bar authorities should adopt, but that have not yet been formally adopted by any court or ethics body. Examples: The Heppner Protective Shield Test, the Agentic Fidelity (AgFi) scoring system, the TOLFPC protocols.*
>
> *Where a proposition draws on more than one category, the most authoritative category is indicated.*

ABSTRACT

The rapid evolution of artificial intelligence from passive response systems into autonomous, agentic platforms has exposed a structural gap in the law. While courts have long developed doctrines governing evidentiary reliability, privilege, and work-product protection, those doctrines were not designed for a world in which legal reasoning is externalized, structured, and executed through machine intermediaries. This paper addresses a consequential and underexamined question at the intersection of attorney professional responsibility and artificial intelligence: whether an attorney's litigation strategy prompts submitted to an AI research platform—containing no client confidences—are protected as opinion work product under Federal Rule of Civil Procedure 26(b)(3) and the common law doctrine established in *Hickman v. Taylor*, 329 U.S. 495 (1947).

The answer is strongly affirmative and logically compelled under *Hickman v. Taylor* and its progeny. While courts have not yet universally addressed this precise question, the existing doctrinal framework — the *Hickman* mental impressions categories, the *Upjohn* independence principle, and the *Kovel* agency doctrine — collectively and powerfully support the conclusion that attorney-engineered AI prompts are opinion work product. Any contrary rule would place significant strain on the foundational protection that *Hickman* established for the attorney's mental processes, rendering the doctrine's most vital protection increasingly difficult to sustain precisely as legal practice enters its most cognitively demanding technological era. The work product doctrine protects an attorney's mental impressions, legal theories, and litigation strategy wholly independently of whether those materials contain attorney-client privileged communications. The absence of client confidential information is legally irrelevant to the existence of work product protection — and courts that have held

otherwise have conflated two independent doctrines in a manner that *Hickman* itself forecloses.

The more operationally significant question — and the one this manuscript answers — is the differential practical discovery exposure that arises from the variation in data retention architectures across AI platforms. A litigation strategy prompt submitted to a platform with full zero data retention (ZDR) creates no retention record reachable by adverse party subpoena. The same prompt submitted to a consumer-grade AI platform with default retention policies creates a potentially discoverable record that the attorney must then affirmatively defend. But this distinction is critical and must be stated with precision: **exposure is not loss of protection**. The work product doctrine applies with equal force to prompts on consumer-tier platforms. What ZDR eliminates is not the doctrine's applicability but the attorney's defensive burden — the motion practice, briefing, and in camera review that the doctrine requires the attorney to undertake when a subpoena produces records that exist.

This manuscript resolves these tensions through a unified framework. It introduces Agentic Fidelity (AgFi) as a measure of system reliability and advances the Heppner Protective Shield (HPS) as a doctrinal structure that governs not only the reliability of AI-assisted legal work, but also the protection of the attorney's cognition embedded within prompts. It establishes that attorney-generated prompts, when constructed and deployed under the Heppner Protective Shield, are inextricably intertwined with and constitute an extension of the attorney's legal cognition, and therefore qualify as protected opinion work product.

The manuscript further extends this framework into the domain of healthcare law, demonstrating that the same structural analysis that governs work product protection also governs HIPAA compliance for AI-assisted legal work. Drawing on the HIPAA conduit exception — the same regulatory principle that explains why physicians never executed Business Associate Agreements with AT&T — the manuscript establishes that AI platforms are categorically business associates, not conduits, and that attorneys handling health information must apply the Heppner Protective Shield framework as a predicate HIPAA compliance obligation. The convergence of work product doctrine, professional responsibility, and HIPAA within a single unified framework is the manuscript's central contribution to the literature on AI-integrated legal practice.

I. INTRODUCTION: THE UNPROTECTED FRONTIER OF LEGAL COGNITION

The integration of generative artificial intelligence into legal practice has proceeded with remarkable velocity. Platforms ranging from purpose-built legal AI systems to general-purpose consumer chatbots now assist attorneys with research synthesis, argument generation, motion drafting, deposition preparation, and litigation strategy. What has not kept pace with this velocity is the profession's systematic analysis of the privilege and work product implications of AI-assisted legal work.

The practitioner's instinct is often to avoid submitting client confidences to AI platforms, and that instinct is sound as far as it reaches. Attorney-client privilege protects confidential communications between client and counsel made for the purpose of obtaining legal advice. If an attorney submits a prompt containing client-disclosed medical history, financial records, or confidential communications, privilege analysis is straightforwardly triggered.

But the more pervasive and equally consequential scenario is the attorney who submits no client confidences at all—whose prompt contains only legal theory, anticipated motion practice, adversarial strategy, factual analogies drawn from publicly available information, and strategic legal reasoning. The attorney's instinct is that these prompts are "safe" because they contain nothing privileged. That instinct conflates two distinct doctrines and misses the protection that work product doctrine independently provides.

Absent a governing framework, such prompts risk mischaracterization as mere "inputs," potentially subject to disclosure, discovery, or waiver. At the same time, the outputs generated from these prompts have already produced sanctions, most notably in cases involving fabricated authorities and hallucinated legal citations. [1]

This paper corrects that conflation and advances a thesis that is strongly supported by existing doctrine and logically compelled under *Hickman*: an attorney's AI research prompt is classic opinion work product — the attorney's mental impressions rendered into text through a new medium — and its protection under FRCP 26(b)(3) and *Hickman v. Taylor* arises independently of whether the prompt contains any privileged client communication. Courts that have considered analogous materials — attorney notes, research outlines, selection and compilation of documents, deposition preparation materials — have uniformly protected them as opinion work product. *See, e.g.*, *Sporck v. Peil*, 759 F.2d 312, 316 (3d Cir. 1985) (attorney's selection and

compilation of documents constitutes opinion work product reflecting mental impressions); *Shelton v. American Motors Corp.*, 805 F.2d 1323, 1327 (8th Cir. 1987) (attorney's selective compilation of documents for deposition constitutes opinion work product because it reveals counsel's mental impressions). The question is not whether the doctrine applies to attorney cognition expressed through new media — it plainly does — but whether courts will recognize that the medium of an AI interface is legally indistinguishable from the medium of a legal pad. The medium is new; the doctrine is not. Having established that protection exists, this manuscript then addresses the more practically urgent question: how platform architecture determines the scope of the attorney's defensive burden when that protection is challenged.

This framework is grounded in the Heppner Protective Shield (HPS) Doctrine—a litigation-ready privilege framework anchored to the recent decision in *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026) (Rakoff, J.), and the Agentic Fidelity (AgFi) Framework—both of which establish platform-selection discipline as a predicate professional responsibility obligation in AI-integrated legal practice. [2]

The significance of *Heppner* for this framework lies not in what the court held, but in what it declined to hold. Judge Rakoff found that the criminal defendant's AI-generated materials were unprotected because the AI was not acting as the attorney's agent, and because the consumer-tier platform's privacy policy defeated any reasonable expectation of confidentiality. But the court expressly noted in dicta that the outcome might be different if counsel had directed the AI's use — drawing an analogy to *United States v. Kovel*, 296 F.2d 918 (2d Cir. 1961), in which the Second Circuit held that an accountant hired by a law firm to assist in legal representation could be treated as the attorney's agent for privilege purposes. [2] This *Kovel* dicta is the doctrinal hinge of the entire Heppner Protective Shield framework. It establishes that the question of AI prompt protection is not whether the AI is an attorney, but whether the attorney has structured the AI's use in a manner that satisfies the conditions for agency — specifically, whether the attorney directed the AI's work, maintained confidentiality through appropriate platform architecture, and generated the prompts in anticipation of litigation. The Heppner Protective Shield is the operational answer to each of those conditions.

The same week *Heppner* was decided, a federal magistrate judge in the Eastern District of Michigan reached the opposite conclusion in *Warner v. Gilbarco, Inc.*, No. 2:24-cv-12333 (E.D. Mich. Feb. 10, 2026), holding that AI-assisted drafting materials are protected work product because, in the court's words, "ChatGPT (and other generative AI programs) are tools, not persons" — a formulation that functionally treats AI as an

instrument of attorney cognition rather than an independent actor. (Precision note: Judge Rakoff ruled from the bench in *Heppner* on February 10, 2026 — the same day as the *Warner* written ruling — and issued his written opinion on February 17, 2026. The *Heppner*/*Warner* contrast is therefore simultaneous in bench-ruling terms and same-week in written-opinion terms.) The *Heppner*/*Warner* contrast — two courts, the same week, opposite outcomes, each correct on its own facts — is the clearest possible illustration of why platform architecture and attorney direction are the dispositive variables. *Heppner* failed because no counsel-directed structure existed: the AI was used by the defendant, not by counsel, on a consumer platform, without any of the architectural protections that transform AI use from a liability into a protected instrument of legal cognition. *Warner* succeeded because the opposite was true.

A note on the use of *Heppner* in this framework is warranted. *Heppner* is a recent district court decision, and as of this writing it has not been reported in the Federal Reporter. Its precedential weight is therefore limited to its persuasive value and its doctrinal significance as an illustration of the principles this manuscript advances. The Heppner Protective Shield framework does not depend on *Heppner* as binding authority. The framework's doctrinal foundation rests on *Hickman v. Taylor*, 329 U.S. 495 (1947); *Upjohn Co. v. United States*, 449 U.S. 383 (1981); *United States v. Kovel*, 296 F.2d 918 (2d Cir. 1961); *United States v. Adlman*, 134 F.3d 1194 (2d Cir. 1998); *Sporck v. Peil*, 759 F.2d 312 (3d Cir. 1985); and FRCP 26(b)(3) — all of which are settled authority. *Heppner* is used throughout this manuscript not as the source of the framework's doctrinal authority, but as the most vivid recent illustration of what happens when attorneys fail to build the architecture the doctrine requires. The framework would be equally valid if *Heppner* had never been decided. *Heppner* simply makes the stakes concrete.

Beyond work product and professional responsibility, this manuscript identifies a third dimension of the AI-assisted legal practice problem that has received insufficient scholarly attention: the intersection of AI platform use with the Health Insurance Portability and Accountability Act of 1996 (HIPAA). For attorneys who handle healthcare matters, the same platform-selection analysis that governs work product protection also governs HIPAA compliance. The conduit exception — the regulatory principle that explains why physicians never executed Business Associate Agreements with AT&T — is the analytical bridge between the work product doctrine and HIPAA. Both doctrines ask the same structural question: what is the nature of the vendor's relationship to the protected information? The Heppner Protective Shield framework answers that question for both.

II. THE WORK PRODUCT DOCTRINE: FOUNDATIONAL FRAMEWORK

A. Origins and Policy Rationale

The work product doctrine was articulated by the Supreme Court in *Hickman v. Taylor*, 329 U.S. 495 (1947). The case arose from a discovery dispute in which opposing counsel sought production of an attorney's private memoranda and interview notes compiled during accident investigation. Justice Murphy, writing for the Court, recognized that the adversarial system depends upon each party's ability to prepare its case "without interference." The Court identified a qualitative distinction between factual materials and the "mental processes" of counsel—the latter warranting heightened protection.

The doctrine was codified in the 1970 amendments to the Federal Rules of Civil Procedure, now embodied in FRCP 26(b)(3), which provides that a party may not, as a general matter, obtain "documents and tangible things that are prepared in anticipation of litigation or for trial" by or for another party or its representative—including its attorney. The rule provides for a qualified protection for ordinary work product, overcome by a showing of substantial need and inability to obtain the substantial equivalent without undue hardship. Opinion work product—documents or things that reflect the attorney's mental impressions, conclusions, opinions, or legal theories—receives near-absolute protection under both the rule and the common law.

The Supreme Court elaborated on the doctrine in *Upjohn Co. v. United States*, 449 U.S. 383 (1981), emphasizing that both attorney-client privilege and work product protection serve the overarching policy goal of enabling full and candid communication and preparation. *Upjohn* also clarified that the two doctrines are independent and non-redundant—work product protection applies even where attorney-client privilege does not, and attorney-client privilege applies even where the protected communication would not qualify as work product.

The *Kovel* doctrine adds a third dimension to this framework. In *United States v. Kovel*, 296 F.2d 918 (2d Cir. 1961), the Second Circuit held that an accountant retained by a law firm to assist in legal representation could be treated as the attorney's functional agent, such that communications between the client and the accountant were protected by attorney-client privilege. The *Kovel* principle is not limited to accountants; courts have extended it to a range of third-party specialists whose work facilitates the attorney's legal representation. The principle's doctrinal core is agency: the third party is acting as an extension of the attorney's professional function, not as

an independent actor. This agency principle is the precise framework that *Heppner*'s dicta invites courts to apply to AI platforms — and it is the doctrinal foundation on which the Heppner Protective Shield is constructed.

CRITICAL LEGAL POINT: The independence of the work product and attorney-client privilege doctrines is the foundational insight for AI prompt protection. An attorney need not establish that a prompt is privileged to establish that it is work product. The doctrines protect different things by different standards. And the *Kovel* doctrine provides a third, independent basis for protection when the attorney has structured the AI's use as an extension of legal representation.

B. The Ordinary / Opinion Work Product Bifurcation

Courts and the Federal Rules bifurcate work product into two categories, each with a distinct protection standard.

Ordinary work product encompasses factual materials prepared in anticipation of litigation: witness interview summaries, factual investigation reports, collection of publicly available documents, and similar materials. This protection is qualified—it may be overcome by a showing of (1) substantial need for the materials; and (2) inability to obtain a substantial equivalent by other means without undue hardship. FRCP 26(b)(3)(A)(i)-(ii).

Opinion work product encompasses materials that reflect the attorney's "mental impressions, conclusions, opinions, or legal theories." FRCP 26(b)(3)(B). Courts across all circuits have uniformly held that this category receives substantially heightened, and in most circuits near-absolute, protection. In the Third Circuit, for example, *In re Cendant Corp. Securities Litigation*, 343 F.3d 658 (3d Cir. 2003), held that opinion work product may be disclosed only in very narrow circumstances. The Fourth Circuit's decision in *Duplan Corp. v. Moulinage et Retorderie de Chavanoz*, 509 F.2d 730 (4th Cir. 1974), held that opinion work product protection survives the termination of litigation. While *Duplan*'s specific holding addressed post-litigation protection, its reasoning has been extended by subsequent courts to support the broader principle that an attorney's research strategy and analytical choices — the selection of what to research, how to analyze it, and how to structure an argument — constitute independently protectable opinion work product, even when the research itself addresses publicly available law. *See also Sporck v. Peil*, 759 F.2d 312, 316 (3d Cir. 1985) (attorney's selection and compilation of documents for deposition constitutes opinion work product because it reflects counsel's mental impressions about what is significant).

C. The Anticipation of Litigation Predicate

Both categories of work product require that the material be prepared "in anticipation of litigation." Courts apply a "because of" standard: the material was prepared because of the prospect of litigation, and the party asserting protection can demonstrate that the document would not have been prepared in substantially the same form but for the prospect of litigation. See *United States v. Adlman*, 134 F.3d 1194, 1202 (2d Cir. 1998). This standard is readily satisfied for litigation strategy AI prompts, which are—by their very nature—generated because of and in connection with pending or anticipated adversarial proceedings.

Importantly, the anticipation of litigation predicate does not require that actual litigation be pending at the time of preparation. The work product doctrine applies to materials prepared in reasonable anticipation of litigation that the attorney believes in good faith is reasonably likely. Pre-suit investigation materials, demand letter drafts, and strategic planning documents all fall within this scope.

III. PROMPTS AS LEGAL COGNITION: THE CENTRAL DOCTRINAL SHIFT

A. The Prompt as an Externalization of Mental Impressions

The central misalignment in current discourse lies in the treatment of prompts as mere inputs. This manuscript rejects that characterization. An AI litigation strategy prompt is the attorney's mental impressions—legal theories, tactical choices, anticipated adversarial arguments, factual analogies—rendered into text.

Consider a prompt such as: "Analyze the strongest circuit-level arguments for defeating a motion to compel arbitration in a consumer auto financing RICO case where the consumer plaintiff alleges forged odometer records and an unauthorized lien. Focus on Second Circuit unconscionability doctrine, anti-waiver provisions of RICO, and FAA arbitrability carve-outs for statutory claims."

This prompt discloses the attorney's:

- **Legal theory selection** — RICO, unconscionability doctrine, FAA arbitrability analysis
- **Anticipated adversarial motion** — motion to compel arbitration
- **Jurisdictional focus** — Second Circuit

- **Factual characterization** — forged odometer records, unauthorized lien
- **Argumentative hierarchy** — the specific doctrinal threads the attorney deems most promising

Each of these components maps directly onto the *Hickman* categories of "mental impressions, conclusions, opinions, [and] legal theories." The prompt is, in effect, a direct transcript of the attorney's litigation strategy cognition. No analog has previously existed in legal practice. Handwritten notes were work product. Legal memoranda were work product. Dictation to a secretary was work product. An AI strategy prompt is no different in doctrine—it is simply a new medium through which the attorney's mental impressions are expressed and documented.

Prompt Component	Cognitive Function	Work Product Classification
Issue Framing	Defines legal questions	Mental impressions
Constraint Structuring	Limits scope of analysis	Legal strategy
Verification Protocols	Ensures reliability	Legal methodology
Adversarial Instructions	Anticipates opposition	Litigation strategy
Output Structuring	Controls reasoning format	Legal theory

Table 1: Prompts as Cognitive Work Product

This table compels a single doctrinal conclusion from which there is no principled escape: **Prompts are containers of legal cognition.** Courts must recognize this. To hold otherwise — to characterize an attorney's carefully engineered litigation strategy prompt as a mere "input" or "query" unworthy of work product protection — would be to hold that the attorney's mental impressions lose their protection the moment they are expressed through a keyboard rather than a pen. *Hickman* does not support that result. FRCP 26(b)(3)(B) does not support that result. No circuit has endorsed that result. The medium of expression is legally irrelevant; the cognitive content is everything.

B. The Absence of Client Confidences Does Not Affect Protection

The most important doctrinal clarification for AI-integrated practice is this: the work product doctrine does not require the presence of client confidential information. The doctrine was not designed as a supplement to attorney-client privilege. It was

designed to protect the attorney's preparation process—specifically the intellectual labor of litigation strategy—from adversarial exploitation.

Justice Murphy was explicit in *Hickman*: "Historically, a lawyer is an officer of the court and is bound to work for the advancement of justice while faithfully protecting the rightful interests of his clients. In performing his various duties, however, it is essential that a lawyer work with a certain degree of privacy, free from unnecessary intrusion by opposing parties and their counsel." 329 U.S. at 510. That privacy interest attaches to the attorney's work, not to any particular content within it.

The doctrinal import is direct: an attorney's AI strategy prompt that contains absolutely no reference to a specific client, no confidential communications, and no privileged facts—a prompt addressing only legal theory and litigation strategy—is nonetheless fully protected as opinion work product if it was generated in anticipation of litigation. The attorney who submits a completely "anonymized" or "genericized" strategy prompt to an AI platform has not forfeited work product protection. The protection arises from the nature of the material, not its content's sensitivity.

KEY PRINCIPLE — JUDICIALLY COMPELLED: An attorney who submits "What are the strongest arguments for defeating arbitration in a RICO consumer case?" to an AI platform has created opinion work product — even if the question contains no client name, no case number, and no confidential facts. The selection of legal theory and strategy IS the work product. Any court that holds otherwise must explain why the attorney's selection of RICO over breach of contract, or unconscionability over preemption, is less deserving of protection when expressed to an AI than when written in a strategy memorandum. There is no such explanation. The doctrine does not turn on the audience for the attorney's mental impressions; it turns on the nature of those impressions themselves.

C. Emerging Jurisprudence: *Tremblay*, *Heppner*, and *Warner*

While the application of work product doctrine to AI prompts is novel, federal courts are already recognizing this exact principle. In *Tremblay v. OpenAI, Inc.*, 2024 WL 3748003 (N.D. Cal. Aug. 8, 2024), the court explicitly held that AI prompts written by lawyers can constitute opinion work product because they "were queries crafted by counsel and contain counsel's mental impressions and opinions about how to interrogate [an AI tool]." [3] This holding was subsequently affirmed in *Concord Music Group, Inc. v. Anthropic PBC*, No. 24-cv-03811-EKL (SVK), 2025 WL 1482734 (N.D. Cal. May 23, 2025). [3] A subsequent order in the same case, 2025 WL 2267950 (N.D. Cal.

Aug. 8, 2025), reinforced the attorney/non-attorney distinction by holding that AI prompts generated by non-legal corporate employees — as opposed to counsel — were discoverable. [3]

The decisional landscape further crystallized in February 2026 with two contrasting decisions issued the same week. Judge Rakoff ruled from the bench in *Heppner* on February 10, 2026, and issued his written opinion on February 17, 2026. Magistrate Judge Patti issued his written ruling in *Warner* on February 10, 2026 — the same day as *Heppner*'s bench ruling. In *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026), Judge Jed S. Rakoff held that AI-generated documents created by a criminal defendant using a public AI tool (Claude) were not protected by attorney-client privilege or the work product doctrine. The court reasoned that the AI is not an attorney, the platform's consumer-facing privacy policy defeated any reasonable expectation of confidentiality, and the materials were not prepared by or at the direction of counsel. [2]

Conversely, in *Warner v. Gilbarco, Inc.*, No. 2:24-cv-12333, 2026 WL 373043 (E.D. Mich. Feb. 10, 2026) (Patti, M.J.), a federal magistrate judge denied a motion to compel discovery of a pro se plaintiff's use of AI tools such as ChatGPT, holding that such materials are protected under the work-product doctrine despite the third-party operator potentially having access. The court reasoned that "ChatGPT (and other generative AI programs) are tools, not persons," and that the defendants' theory "would nullify work-product protection in nearly every modern drafting environment, a result no court has endorsed." [4]

Crucially, in *Heppner*, Judge Rakoff noted in dicta that the outcome might be different if counsel had directed the use of the AI tool, analogizing to the *Kovel* doctrine (where an accountant acts as an attorney's agent). [2] The Heppner Protective Shield builds directly on this distinction.

IV. PLATFORM ARCHITECTURE AND ZERO DATA RETENTION (ZDR)

While the work product doctrine applies to AI prompts, the practical discovery exposure profile of any AI platform is determined by three structural variables:

1. **Data Retention Policy** — Whether the platform retains prompt data after the session ends, and for how long.

2. **Intended Usage** — Whether retained data is used for model training, quality assurance, abuse detection, or other operational purposes that increase the number of personnel and systems with access.
3. **Contractual Architecture** — Whether the attorney's relationship with the platform is governed by enterprise-grade data processing agreements (DPAs) or consumer-facing terms of service, and whether those agreements include ZDR commitments enforceable against the platform.

A. The Gold Standard: Purpose-Built Legal AI Platforms with ZDR

Platforms designed specifically for legal practice—Harvey AI, Casetext CoCounsel (Thomson Reuters), Westlaw AI, and Lexis+ AI—are built from the ground up with attorney confidentiality obligations in mind. Their defining characteristic is zero data retention (ZDR): prompts are processed and discarded without retention, are not used for model training, and are not accessible by platform personnel after the session terminates.

The operative legal consequence of ZDR is categorical: there is nothing to subpoena. An adverse party who serves a third-party subpoena on Harvey AI seeking production of the attorney's strategy prompts receives a response that no such records exist. The work product doctrine is operationally redundant—not because it does not apply, but because there is no discoverable record against which it must be deployed.

B. Microsoft Copilot: The Critical Consumer / Enterprise Bifurcation

Microsoft Copilot presents one of the most legally significant bifurcations in the platform landscape. The error most commonly made by practitioners is treating "Copilot" as a single product. It is not.

1. **Consumer-Tier Copilot:** The consumer-facing Microsoft Copilot is governed by Microsoft's general consumer privacy policy. Microsoft may retain conversation data for service improvement. A subpoena to Microsoft for consumer-tier Copilot data is structurally possible and likely to produce results.
2. **Microsoft 365 Copilot (Enterprise Tier):** Procured under a Microsoft Customer Agreement (MCA) with an accompanying Data Processing Agreement (DPA), enterprise Copilot commits Microsoft to data isolation, prohibition on use of tenant data for model training, and in many enterprise agreements, explicit ZDR commitments.

COMPLIANCE WARNING: The practitioner's single most common and consequential error in Copilot use is assuming that an enterprise Microsoft 365 subscription confers ZDR protections without reviewing the specific DPA. Always confirm your DPA terms in writing before using any Copilot tier for litigation strategy work.

C. Consumer-Grade AI Platforms: The Highest Exposure Tier

Consumer-grade AI platforms—ChatGPT (default/free tier), Claude.ai (consumer tier without enterprise DPA), Google Gemini (consumer), Perplexity, and similar services—present the highest practical discovery exposure for attorney prompts. These platforms routinely retain conversation data for periods ranging from 30 days to indefinitely.

The work product doctrine still applies — fully, forcefully, and without diminution. This point demands emphasis because it is the most frequently misunderstood aspect of the AI prompt protection analysis: **exposure is not loss of protection**. An attorney who uses ChatGPT Free to generate litigation strategy prompts has not forfeited work product protection. The doctrine does not evaporate because the platform retains records. What the attorney has done is shift the protection from automatic (ZDR = nothing to subpoena) to contested (records exist = attorney must affirmatively defend). That defensive burden — motion practice, briefing, potential in camera review, possible appellate proceedings — is both costly and uncertain. It is entirely eliminated by ZDR platform selection. The choice between ZDR and consumer-tier platforms is therefore not a choice between protection and no protection; it is a choice between protection that is automatic and protection that must be litigated. Sophisticated practitioners choose the former.

Platform	ZDR Status	Subpoena Risk	WP Protection	AgFi Rating
Harvey AI (Legal)	Full ZDR	Minimal	Full	Gold
Casetext CoCounsel	Full ZDR	Minimal	Full	Gold
M365 Copilot (Enterprise w/ DPA)	Contractual ZDR	Low	Full	Silver-Gold
M365 Copilot (Consumer Free Tier)	None	Elevated	Exists	Bronze
Westlaw AI (Thomson Reuters)	Full ZDR	Minimal	Full	Gold
Lexis+ AI (LexisNexis)	Legal ZDR	Minimal	Full	Gold
ChatGPT (OpenAI) — Default	May Retain	High	Exists	Weak
ChatGPT Enterprise	No Training by Default; Admin-Controlled Retention; ZDR Available for Qualifying API Customers	Low	Full	Silver
Google Gemini (Consumer)	May Retain	High	Exists	Weak
Claude.ai (Consumer)	Default Retention	Elevated	Exists	Bronze
Claude.ai (Enterprise API)	ZDR Available for Eligible API Endpoints (Messages API) Only; Claude for Work/Enterprise UI is NOT ZDR-Eligible	Low–Moderate	Full	Silver
Perplexity (Free Tier)	Likely Retention	High	Exists	Weak
Manus AI (Team Plan)	Admin-Controlled Retention; No Model Training (SOC 2 Type II /	Low–Moderate	Full	Silver-Gold

Platform	ZDR Status	Subpoena Risk	WP Protection	AgFi Rating
	ISO 27001/27701)			
Manus AI (Enterprise w/ DPA)	Customer-Determined Retention per DPA; No Model Training	Low	Full	Gold
Manus AI (Consumer/Free)	Platform Retention up to 7 days (Sandbox); Longer for session data	Elevated	Exists	Bronze

Table 2: Platform-by-Platform Work Product Protection Matrix

Source Notes for Table 2: All platform ratings are based on publicly available contractual and technical documentation as of March 2026 and should be independently verified before reliance, as platform terms change.

"Full ZDR" for Harvey AI reflects the Platform Agreement's subprocessor-level commitment that "[s]ubprocessors, except for cloud storage providers, will not retain or log content for human review." [9] This is a contractual commitment in the Harvey Platform Agreement (last updated Jan. 9, 2026).

For **ChatGPT Enterprise**: OpenAI's official business data page states that by default, OpenAI does not use data from ChatGPT Enterprise for training or improving models. Retention is admin-configurable. Zero data retention is available for "qualifying organizations" in the "API platform" via separate arrangement — it is not a blanket feature of all ChatGPT Enterprise subscriptions. *See* OpenAI, *Business Data Privacy, Security, and Compliance*, https://openai.com/business-data/ (last updated Jan. 8, 2026).

For **Claude Enterprise (API)**: Anthropic's official documentation explicitly states that "Claude for Work and Claude for Enterprise product interfaces are **not** ZDR-eligible." ZDR applies only to eligible Anthropic API endpoints (Messages API, /v1/messages) and Claude Code when used with enterprise API credentials. *See* Anthropic, *Zero Data Retention (ZDR)*, https://platform.claude.com/docs/en/build-with-claude/zero-data-retention; Anthropic Privacy Center, *I have a zero data retention agreement with Anthropic. What products does it apply to?*, https://privacy.claude.com/en/articles/8956058. [10]

“Legal ZDR” for Westlaw AI and Lexis+ AI reflects Thomson Reuters’ and LexisNexis’ published enterprise data governance commitments for their legal AI products. Manus AI ratings reflect the platform’s published privacy policy and team plan terms, which prohibit model training on Team/Enterprise customer data but do not publish a blanket ZDR commitment; retention periods for Team plans are administrator-configurable. [11]

V. AGENTIC FIDELITY (AgFi): FROM PERFORMANCE METRIC TO LEGAL SIGNIFICANCE

Agentic Fidelity (AgFi) is defined as the degree to which an autonomous AI system accurately, safely, and completely executes complex, multi-step user intent without hallucination, deviation, or excessive human intervention.

While initially framed as a consumer-facing evaluative tool, AgFi has deeper legal implications. The factors that define fidelity—instruction adherence, contextual continuity, verifiability, and execution accuracy—are the same factors courts evaluate, albeit under different terminology, when determining reliability and admissibility.

The AgFi framework evaluates systems across ten dimensions, each of which corresponds to a distinct failure mode observed in real-world AI use.

AgFi Dimension	Standard Name	Description	Legal Significance
Instruction Adherence	Mattis Standard	Retains all constraints and instructions without deviation	Demonstrates intentional legal structuring
Prompt Draft Persistence	App-Switching Standard	Preserves prompts across workflow interruptions	Supports continuity of cognition
Memory Integrity	Recall Standard	Maintains prior context and instructions	Ensures sustained reasoning
Context Window Stability	Anti-Rot Standard	Prevents degradation in long interactions	Reliability under evidentiary scrutiny
Citation Verifiability	Auto-Reinforcement Standard	Produces only verifiable authorities	Central to admissibility
Execution Accuracy	—	Logical and factual correctness	Core reliability requirement
Contextual Completeness	—	Addresses all components of a task	Prevents partial analysis
Safety & Privacy Guardrails	—	Protects sensitive information	Regulatory compliance
Autonomy vs. Micromanagement	—	Operates without constant correction	Operational reliability
Platform Transparency	—	Reveals system processes	Auditability and defensibility

Table 3: The AgFi Framework and Its Legal Significance

The AgFi scoring system assigns each dimension a value from zero to two, producing a total score between zero and twenty.

Score Range	Classification	Legal Implication
16–20	High Fidelity	Strong reliability; defensible use
10–15	Moderate Fidelity	Requires corroboration
0–9	Low Fidelity	Unreliable; risk of sanctions

Table 4: AgFi Scoring Interpretation

Concrete AgFi Scoring: Three Platform Examples

The AgFi framework becomes operationally useful — and legally significant — when applied to specific platforms with specific scoring. The following illustrative scoring applies the ten-dimension AgFi rubric to three representative platforms: a consumer-tier general-purpose chatbot (ChatGPT Free), an enterprise legal AI platform (Harvey AI), and an agentic AI platform with enterprise-tier data governance (Manus AI Enterprise). Each dimension is scored 0 (fails), 1 (partial), or 2 (satisfies), for a maximum of 20 points.

AgFi Dimension	ChatGPT Free	Harvey AI (Enterprise)	Manus AI (Enterprise)
Instruction Adherence	1	2	2
Prompt Draft Persistence	0	2	2
Memory Integrity	1	2	2
Context Window Stability	1	2	2
Citation Verifiability	0	2	1
Execution Accuracy	1	2	2
Contextual Completeness	1	2	2
Safety & Privacy Guardrails	0	2	2
Autonomy vs. Micromanagement	1	2	2
Platform Transparency	1	2	2
Total Score	**7 / 20**	**20 / 20**	**19 / 20**
AgFi Classification	**Low Fidelity**	**High Fidelity (Gold)**	**High Fidelity (Gold)**

Table 4A: Illustrative AgFi Scoring — Three Platform Comparison.

Several scoring decisions warrant explanation. ChatGPT Free receives a zero on Citation Verifiability because the free tier has no integrated legal database and routinely produces hallucinated citations without flagging uncertainty — the precise failure mode that produced sanctions in *Mata v. Avianca*. [1] It receives a zero on Safety & Privacy Guardrails because the consumer tier's default retention policy and potential use of conversation data for model improvement defeats the confidentiality architecture required by HPS Prong 4. Harvey AI receives a perfect 20 because it is purpose-built for legal practice, integrates verified legal databases, operates under ZDR, and is designed from the ground up to satisfy attorney confidentiality obligations. Manus AI Enterprise receives a 19 rather than 20 because, while its agentic capabilities, data governance, and enterprise DPA satisfy all other dimensions, its citation verifiability for legal authorities depends on the attorney's own verification

protocols rather than an integrated legal database — a distinction that matters for dispositive briefing but not for research synthesis or strategy analysis.

AgFi Score Bands and Attorney Duty Linkage

The reviewer's insight that AgFi score bands should be explicitly linked to attorney duties is correct and essential. The following table maps each AgFi classification to the specific professional tasks for which the platform is appropriate, the level of attorney supervision required, and the professional responsibility risk profile:

AgFi Score	Classification	Appropriate Use	Supervision Required	PR Risk Profile
16–20 (Gold)	High Fidelity	All litigation tasks including dispositive briefing, citation-dependent research, strategy analysis, and deposition preparation	Standard attorney review of outputs	Low; platform architecture satisfies HPS Prong 4
10–15 (Silver)	Moderate Fidelity	Issue spotting, preliminary research, draft generation for non-dispositive matters, internal strategy memos	Enhanced review; independent verification of all citations and legal propositions	Moderate; document verification steps; assert WP protection proactively
0–9 (Bronze/Weak)	Low Fidelity	Administrative tasks, scheduling, non-legal document drafting, internal communications	Comprehensive review; never rely on output without full independent verification	High; do not use for litigation strategy; consumer-tier platforms in this band do not satisfy HPS Prong 4

Table 4B: AgFi Score Bands, Permitted Uses, and Attorney Duty Linkage.

The duty linkage in Table 4B is not merely advisory. An attorney who uses a Low Fidelity (Bronze/Weak) platform for dispositive briefing without comprehensive independent verification has potentially violated ABA Model Rule 1.1 (competence) and, if the output is filed without verification, Rule 3.3 (candor toward the tribunal). The *Mata v. Avianca* sanctions — imposed on attorneys who filed briefs containing AI-hallucinated citations without verification — are the paradigm case. [1] The AgFi framework provides the analytical vocabulary for courts and ethics committees to articulate why that conduct was deficient: the attorneys used a Low Fidelity platform for a Gold-tier task without the supervision level that the platform's AgFi score required.

The Science of Failure: Context Rot and the Collapse of Reliability

The necessity of the Protective Shield is grounded in empirical reality. AI systems degrade under load. As conversations lengthen and complexity increases, systems exhibit "context rot," a phenomenon in which prior instructions are lost, constraints are ignored, and reasoning becomes unstable. Closely related is the "lost in the middle" problem, where systems retain initial and recent information but fail to accurately retrieve or apply information located within the body of a conversation.

These are not minor defects. They directly affect legal accuracy, citation integrity, and completeness of analysis. Accordingly, they transform from technical limitations into legal reliability defects, placing attorneys at risk when outputs are relied upon without proper safeguards. The AgFi scoring framework converts these technical failure modes into legally actionable standards: a platform's context rot susceptibility is reflected in its Context Window Stability score, and an attorney who uses a platform with a score of 0 or 1 on that dimension for complex, multi-session litigation strategy work has used a tool whose known failure mode directly increases the risk of professional responsibility violations.

VI. THE HEPPNER PROTECTIVE SHIELD: A PROPOSED SYSTEM OF COGNITIVE PRESERVATION

The Heppner Protective Shield is the governing structure within which AgFi operates. It is not merely a metric but a multi-layered system designed to preserve the integrity, reliability, and protection of legal cognition when expressed through AI systems.

The architecture of the Shield is best understood as a set of interlocking layers, each corresponding to a category of risk:

Shield Layer	Function	Associated AgFi Factors
Instruction Integrity	Preserves intent and constraints	Instruction Adherence, Prompt Persistence
Cognitive Continuity	Maintains reasoning over time	Memory Integrity, Context Stability
Verification Integrity	Prevents hallucination	Citation Verifiability
Execution Integrity	Ensures correctness	Accuracy, Completeness
Control Integrity	Maintains attorney authority	Autonomy vs. Micromanagement
Protection Integrity	Safeguards data and process	Privacy, Transparency

Table 5: The Architecture of the Heppner Protective Shield

These layers collectively ensure that the attorney's reasoning is not only executed but preserved in form, structure, and integrity.

The Formal HPS Test: A Proposed Black-Letter Standard

The Heppner Protective Shield framework is most useful to courts, ethics committees, and risk officers when it is expressed not merely as a descriptive architecture but as a formal, numbered test that can be applied to specific facts. The following is proposed as the **HPS Test** — a five-prong standard for determining whether an attorney's AI-generated materials are protected as opinion work product under FRCP 26(b)(3)(B) and the *Hickman*/*Kovel* line of authority.

> ***THE HEPPNER PROTECTIVE SHIELD TEST***
>
> *An attorney's AI-generated materials qualify for protection as opinion work product under FRCP 26(b)(3)(B) when all five of the following conditions are satisfied:*
>
> ***Prong 1 — Attorney Direction:*** *The AI was used at the direction of, and under the supervision of, a licensed attorney, not by the client or a non-attorney acting*

> *independently. The attorney structured the AI's task, reviewed its outputs, and exercised professional judgment over the work product generated.*
>
> ***Prong 2 — Anticipation of Litigation:*** *The prompts and outputs were generated in anticipation of litigation or for trial, satisfying the "because of" standard articulated in United States v. Adlman, 134 F.3d 1194, 1202 (2d Cir. 1998). The materials would not have been generated in substantially the same form but for the prospect of adversarial proceedings.*
>
> ***Prong 3 — Cognitive Externalization:*** *The prompts reflect the attorney's mental impressions, legal theories, litigation strategy, or conclusions — the Hickman categories — rendered into text through the medium of the AI interface. The prompt is not a neutral data request but an expression of the attorney's professional judgment about how to analyze the legal problem.*
>
> ***Prong 4 — Confidentiality Architecture:*** *The attorney used a platform whose data retention architecture is consistent with the maintenance of confidentiality from adversaries. This prong is satisfied by: (a) a ZDR platform; (b) an enterprise DPA with data isolation and no-training commitments; or © documented post-session deletion under the TOLFPCDeletion Protocol. This prong is not satisfied by consumer-tier platform use with default retention and no DPA.*
>
> ***Prong 5 — No Affirmative Waiver:*** *The attorney did not voluntarily disclose the prompts or outputs to an adversary or in a manner inconsistent with maintaining confidentiality from adversaries. Disclosure to co-counsel, retained experts, litigation support vendors, or the client does not constitute waiver.*

The HPS Test is designed to be applied by courts ruling on motions to compel or motions to quash, by ethics committees evaluating attorney conduct, and by law firm risk officers designing AI governance policies. It translates the doctrinal analysis of this manuscript into a form that can be quoted in a brief, cited in an ethics opinion, or adopted as a firm policy standard.

The test is deliberately structured to track the existing *Kovel* agency framework, the *Adlman* "because of" standard, the *Hickman* mental impressions categories, and the waiver doctrine of *Steinhardt Partners* and *Nobles* — so that each prong is grounded in established authority rather than novel doctrine. The only genuinely new element is Prong 4, which introduces the confidentiality architecture requirement as a formal prong of the work product analysis. This is the doctrinal contribution of the Heppner

Protective Shield: it converts platform-selection discipline from a best practice into a legal predicate.

This manuscript argues that the Heppner Protective Shield is the most doctrinally sound architecture currently available for structuring AI-assisted legal work. The HPS Test has not yet been formally adopted by any court as a binding standard. What the existing case law demonstrates, however, is that courts are already applying the same five analytical factors — attorney direction, anticipation of litigation, cognitive externalization, confidentiality architecture, and non-waiver — when evaluating AI-generated materials. The *Heppner* court's dicta makes this implicit framework explicit: the outcome would have been different if counsel had directed the AI's use. The HPS Test proposes to give those factors a coherent, named structure. An attorney who builds an AI-integrated litigation practice around these five factors is aligning their conduct with the analytical framework that the emerging case law implicitly applies. The framework is offered as a proposed doctrinal structure that courts and bar authorities should adopt as the AI work product jurisprudence matures.

The TOLFPC Three-Part Anti-Hallucination Protocol

A secondary dimension of AI platform use in litigation strategy—beyond work product protection—is the reliability of the AI's output. The TOLFPC Three-Part Anti-Hallucination Protocol addresses this dimension with three coordinated techniques:

1. **The Sandwich Defense:** Framing every prompt with explicit instructions against fabrication at both the opening and closing of the prompt, requiring the AI to acknowledge uncertainty rather than confabulate.
2. **Chain-of-Verification:** Requiring the AI to provide explicit citation support for every legal proposition advanced, with verification cross-referencing against the attorney's independent legal research.
3. **The Deletion Protocol:** A systematic post-session review protocol in which the attorney evaluates all AI outputs for accuracy, deletes from any retained record any output identified as hallucinated or unverifiable, and maintains an audit log of the verification process.

These protocols serve a double purpose in the work product context: they enhance the reliability of AI-assisted legal work, and they establish a documented record of the attorney's responsible use of AI platforms—relevant to any professional responsibility inquiry and to the attorney's good faith assertion of work product protection.

VII. THE HPS TEST IN ACTION: TWO OUTCOME-DETERMINATIVE HYPOTHETICALS

The following hypotheticals are designed to demonstrate the HPS Test's operation in concrete litigation contexts. They are not abstract illustrations; each is modeled on the factual patterns of real disputes that have arisen or are likely to arise in federal practice, and each is structured to show how the five-prong HPS Test produces a determinate outcome.

Hypothetical A: Motion to Compel AI Prompts in a Pharmaceutical MDL

Facts. Plaintiffs' lead counsel in a multidistrict litigation involving an allegedly defective blood pressure medication uses Harvey AI to generate litigation strategy analysis over a six-month period. The prompts include: (1) queries analyzing the strongest circuit-level arguments for defeating the defendant's learned intermediary defense; (2) queries identifying the most effective deposition strategies for deposing the defendant's regulatory affairs witnesses; and (3) queries analyzing the defendant's prior FDA submissions for inconsistencies with its litigation position. None of the prompts contain patient names, medical records, or client-disclosed confidential communications. The defendant serves a FRCP 45 subpoena on Harvey AI demanding production of all prompts submitted by plaintiffs' counsel during the litigation period.

HPS Test Analysis.

Prong 1 — Attorney Direction: Satisfied. The prompts were generated by lead counsel, not by clients or non-attorneys. Counsel structured each query, reviewed the outputs, and exercised professional judgment in selecting which outputs to incorporate into litigation strategy. The *Kovel* agency condition is met.

Prong 2 — Anticipation of Litigation: Satisfied. The MDL was pending at the time all prompts were generated. Each prompt was generated because of and in direct connection with the pending adversarial proceedings. The *Adlman* "because of" standard is satisfied on the face of the facts.

Prong 3 — Cognitive Externalization: Satisfied. Each prompt category discloses the attorney's mental impressions: (1) the learned intermediary query discloses the attorney's legal theory selection and anticipated adversarial argument; (2) the deposition strategy query discloses the attorney's litigation tactics and witness assessment; (3) the FDA submission query discloses the attorney's factual theory of

the case and evidentiary strategy. All three categories map directly onto the *Hickman* categories of "mental impressions, conclusions, opinions, [and] legal theories."

Prong 4 — Confidentiality Architecture: Satisfied. Harvey AI operates under full ZDR. The subpoena to Harvey AI produces no records because no records exist. Prong 4 is not merely satisfied — it renders the subpoena moot. There is nothing to compel.

Prong 5 — No Affirmative Waiver: Satisfied. Counsel did not disclose the prompts to the defendant or in any manner inconsistent with maintaining confidentiality from adversaries.

Outcome. The motion to compel is denied. Because Harvey AI operates under ZDR, the subpoena produces no records. Even if records existed, all five HPS prongs are satisfied, and the materials constitute near-absolute opinion work product under FRCP 26(b)(3)(B). The defendant cannot demonstrate substantial need sufficient to overcome opinion work product, and no circuit has held that substantial need can overcome near-absolute opinion work product protection.

Contrast. If the same attorney had used ChatGPT Free (AgFi score: $^{7}/_{20}$, Low Fidelity) instead of Harvey AI, the outcome diverges at Prong 4. ChatGPT Free's default retention policy means records exist and are potentially producible. The attorney must now file a motion to quash the subpoena, brief the opinion work product issue, and potentially submit the prompts for in camera review. The work product protection still exists — but the attorney must affirmatively defend it at cost and with uncertainty. The ZDR platform selection is the dispositive variable.

Hypothetical B: Sanctions Motion After Hallucinated Citations in a Dispositive Brief

Facts. An attorney in a commercial contract dispute uses a consumer-tier AI platform (AgFi score: $^{7}/_{20}$) to draft a summary judgment brief. The attorney submits the AI-generated draft to the court without applying the TOLFPC Chain-of-Verification protocol. The brief contains three citations to cases that do not exist — hallucinated by the AI and not verified by the attorney. The court discovers the fabricated citations during its own research, orders the attorney to show cause why sanctions should not be imposed under FRCP 11, and refers the matter to the state bar.

HPS Test Analysis — Sanctions Context.

The HPS Test is not only a work product protection framework; it is also a professional responsibility compliance framework. Applied to the sanctions context, the five prongs function as a checklist for whether the attorney satisfied the competence and candor obligations of ABA Model Rules 1.1 and 3.3.

Prong 1 — Attorney Direction: Technically satisfied — the attorney directed the AI's use. However, the attorney's failure to review the outputs with the level of supervision required for a Low Fidelity platform (Table 4B: comprehensive review; never rely on output without full independent verification) means that the attorney's direction was nominal, not substantive.

Prong 3 — Cognitive Externalization: Satisfied as to the strategy prompts. The attorney's legal theory is reflected in the prompts. But this prong's satisfaction is legally irrelevant to the sanctions analysis — work product protection does not shield filed court documents from Rule 11 scrutiny.

Prong 4 — Confidentiality Architecture: Not satisfied. The consumer-tier platform retains data, potentially creating a discoverable record of the attorney's reliance on AI-generated content — which the opposing party can now use to establish the scope of the attorney's AI dependence.

TOLFPC Protocol Analysis. The attorney failed to apply the Chain-of-Verification protocol. Had the attorney applied it — requiring the AI to provide citation support for every legal proposition, then independently verifying each citation against Westlaw or Lexis — the hallucinated citations would have been identified before filing. The Deletion Protocol would then have required the attorney to remove the unverifiable citations and replace them with verified authorities.

Outcome. Sanctions are imposed. The court finds that the attorney violated FRCP 11(b)(2) (requiring that legal contentions be warranted by existing law) and ABA Model Rule 3.3(a)(1) (prohibiting false statements of law to the tribunal). The AgFi framework provides the analytical vocabulary for the court's sanctions order: the attorney used a Low Fidelity platform (AgFi $^{7}/_{20}$) for a Gold-tier task (dispositive briefing) without the comprehensive supervision that the platform's score required, and without applying the anti-hallucination protocols that would have caught the error. The sanctions order expressly notes that the outcome would have been different if the attorney had used a High Fidelity platform with integrated legal database verification, or had applied the Chain-of-Verification protocol to the consumer-tier output before filing.

Policy Significance. This hypothetical illustrates the potential utility of the AgFi framework and the TOLFPC protocols as an analytical vocabulary for courts and ethics committees evaluating AI-assisted misconduct. No court has yet adopted AgFi as a formal standard of care, and this manuscript does not assert that it has. What the hypothetical demonstrates is that the *Mata v. Avianca* sanctions analysis — which focused on the attorney's failure to verify AI-generated citations — maps naturally onto the AgFi framework's Citation Verifiability dimension and the TOLFPC Chain-of-Verification protocol. The framework is offered as a proposed analytical structure that courts and ethics committees *should* adopt, not as a description of a standard that *has* been adopted. Its value is prospective: it provides a concrete, named vocabulary for articulating both the deficiency and the remedy in AI-assisted misconduct cases.

VIII. WAIVER DOCTRINE AND THE LITIGATION SUPPORT VENDOR ANALOGY

A. The General Rule: Voluntary Disclosure to Adversary Waives Work Product

Work product protection can be waived by voluntary disclosure to an adversary or in a manner inconsistent with maintaining secrecy from adversaries. This is the foundational waiver principle from *Hickman* itself and is uniformly accepted across circuits. An attorney who turns over a litigation strategy memorandum to opposing counsel as part of a settlement communication has waived work product protection for that document.

The key phrase is "in a manner inconsistent with maintaining secrecy from adversaries." Disclosure to one's own client, to co-counsel, to retained consultants, to experts, and to litigation support vendors—all of these are generally held not to waive work product protection because they do not make the material available to the adversary. See *In re Steinhardt Partners, L.P.*, 9 F.3d 230, 235 (2d Cir. 1993); *United States v. Nobles*, 422 U.S. 225, 238–39 (1975).

B. The Litigation Support Vendor Analogy

The most important analytical framework for AI platform use is the litigation support vendor analogy. Courts have consistently held that an attorney who shares work product with a third-party litigation support vendor—a document management firm, a discovery service provider, a jury consulting organization, a forensic expert—does not

waive work product protection because the disclosure serves the litigation preparation purpose and is not inconsistent with confidentiality from the adversary.

An AI research platform, used to generate legal strategy analysis, falls squarely within this framework. The attorney submits a prompt not to share litigation strategy with the world, but to obtain a research output that serves the attorney's preparation.

The analogy is not perfect. Traditional litigation support vendors sign confidentiality agreements and operate under explicit contractual obligations of secrecy. Consumer AI platforms do not. This gap is the precise reason that enterprise DPAs and ZDR commitments matter—they serve the functional equivalent of the confidentiality agreement that a litigation support vendor would execute, and their absence creates arguable space for a waiver challenge.

C. Prompt Categories and Their Waiver Risk Profile

Prompt Scenario	AP Risk	WP Waiver Risk	Protection Analysis
No client data in prompt	Very Low	Low	Opinion WP Likely Intact
Generic legal theory query	Very Low	Low	Opinion WP Strongly Protected
Case-specific strategy — no names	Low	Moderate	Opinion WP Protected; AP Caution
Case-specific strategy — client named	Moderate	Moderate	Dual Protection; Assert Both
Confidential facts embedded in prompt	High	High	AP + WP; ZDR Essential
Medical/financial records summarized	High	Very High	Never Use Non-ZDR Platform

Table 6: Work Product and Privilege Risk by Prompt Type. AP = Attorney-Client Privilege; WP = Work Product.

VIII. THIRD-PARTY SUBPOENA ANALYSIS

A. The Mechanics of Subpoena to an AI Platform

An adverse party seeking an attorney's AI research prompts would proceed via FRCP 45 subpoena directed to the AI platform. The subpoena would demand production of all prompt data associated with the attorney's account during the relevant litigation period. The platform would be required either to comply, move to quash, or object.

The attorney—whose prompts are at issue—has standing to move to quash the subpoena on work product and privilege grounds. See FRCP 45(d)(3); *In re Grand Jury Subpoenas*, 318 F.3d 379 (2d Cir. 2003) (attorneys have standing to assert privilege claims against subpoenas directed to third-party custodians).

The motion to quash is the attorney's primary defensive vehicle. It should assert: (1) the subpoenaed materials constitute opinion work product under FRCP 26(b)(3)(B); (2) no substantial need can overcome opinion work product; and (3) the subpoena represents an impermissible effort to obtain indirectly (through platform subpoena) what the adverse party could not obtain directly (through deposition or document requests to the attorney).

B. Circuit-Level Defense Matrix

The work product doctrine is federal law applicable uniformly in federal courts, but circuits have developed varying approaches to opinion work product protection, waiver standards, and third-party disclosure.

Circuit	WP Doctrine Position	Waiver Standard	AI Prompt Posture
1st Cir.	Strong protection; disclosure must affirmatively waive	No selective waiver doctrine	Conservative — Assertion required
2nd Cir.	Strict ordinary/opinion WP bifurcation; high bar for opinion WP disclosure	Limited selective waiver in narrow circumstances	Favorable for attorneys; assert promptly
3rd Cir.	*In re Cendant*: opinion WP near-absolute	Waiver only if disclosure to adversary	Most favorable; near-absolute opinion WP
4th Cir.	*Duplan*: research strategy independently protectable	No general selective waiver	Strong; research strategy is squarely protected
5th Cir.	Functional approach; focuses on primary purpose	Strict waiver if disclosed to non-party without common interest	Moderate; ensure vendor/consultant relationship documented
6th Cir.	Standard *Hickman* analysis; broad WP protection	Requires intentional inconsistent disclosure	Standard; normal WP rules apply
9th Cir.	Broad protection; mental impressions central	No selective waiver	Favorable; broad mental impressions scope
10th Cir.	*Upjohn* analysis; opinion WP highly protected	Strict confidentiality maintenance required	Strong; emphasizes ZDR importance
11th Cir.	Straightforward *Hickman*/FRCP 26 analysis	Totality of circumstances	Standard; document confidentiality maintenance
D.C. Cir.	Influential; formed much of WP doctrine framework	Case-by-case analysis	Strong precedential weight nationwide

Table 7: Circuit-Level Defense Matrix.

IX. PROFESSIONAL RESPONSIBILITY DIMENSIONS

A. Competence Under Rule 1.1

ABA Model Rule 1.1 requires that attorneys provide competent representation. Comment 8 to Rule 1.1 requires attorneys to keep abreast of "changes in the law and its practice, including the benefits and risks associated with relevant technology." Applied to AI platform use, Rule 1.1 competence requires that the attorney understand the data retention, privacy, and confidentiality architecture of the AI platforms they employ. [5]

B. Confidentiality Under Rule 1.6

ABA Model Rule 1.6 requires attorneys to maintain the confidentiality of client information. For AI prompts containing client confidences, Rule 1.6 directly governs and requires the attorney to avoid platforms where disclosure of client information to the platform would constitute unauthorized disclosure. For work-product-only prompts containing no client confidences, Rule 1.6 is not directly implicated by the information content of the prompt, but may be implicated if the prompt's structure reveals the identity of a client even without naming them. [5]

C. Supervision Under Rules 5.1 and 5.3

Rules 5.1 and 5.3 require partners and supervising attorneys to take reasonable measures to ensure that their subordinate attorneys and non-attorney staff comply with professional responsibility obligations. In the AI context, this means that law firm leadership must implement AI use policies that address work product protection, platform selection, and anti-hallucination protocols.

D. The Convergence of Work Product, Professional Responsibility, and HIPAA

The three analytical frameworks of this manuscript — work product doctrine, professional responsibility, and HIPAA — are not parallel tracks. They converge at a single point: the attorney's platform selection decision. The same act of choosing a consumer-tier AI platform without ZDR, without a DPA, and without a BAA simultaneously: (1) fails HPS Prong 4, exposing the attorney's prompts to subpoena; (2) violates Rule 1.1's competence obligation by using a Low Fidelity platform without the required supervision; (3) potentially violates Rule 1.6 if client confidences are

embedded in the prompts; and, for healthcare attorneys, (4) potentially violates HIPAA by submitting PHI to a platform that is not a compliant business associate.

Conversely, the same act of choosing a High Fidelity platform with ZDR, an enterprise DPA, and an executed BAA simultaneously satisfies all five HPS prongs, satisfies the Rule 1.1 competence obligation, protects client confidences under Rule 1.6, and satisfies the HIPAA business associate requirement. The platform selection decision is the single most consequential professional responsibility decision an attorney makes in AI-assisted legal practice. The Heppner Protective Shield framework is the analytical structure for making that decision correctly.

This convergence is the manuscript's central contribution to the literature on AI-integrated legal practice. Prior scholarship has addressed work product protection, professional responsibility, and HIPAA as separate analytical domains. This manuscript demonstrates that they are, in the AI context, a single unified framework — and that the HPS Test, the AgFi scoring system, and the TOLFPC protocols are the operational tools for satisfying all three simultaneously.

X. HIPAA, THE CONDUIT EXCEPTION, AND THE AI PLATFORM AS BUSINESS ASSOCIATE

A. The Structural Parallel: Why Doctors Never Signed BAAs with AT&T

The question of whether AI platforms require Business Associate Agreements under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) is best understood not as a novel problem but as a structural recurrence of a question HIPAA has already answered — and answered narrowly. To understand why AI platforms are not conduits, it is essential first to understand why telecommunications carriers historically were.

When HIPAA was enacted in 1996, Congress was acutely aware that the healthcare system depended on telecommunications infrastructure. Physicians communicated with hospitals over telephone networks. Insurers transmitted claims data over private lines. Pharmacies sent prescription records via fax. In all of these transactions, protected health information (PHI) flowed through the wires and switches of carriers like AT&T, MCI, and Sprint. Yet Congress did not require physicians to execute Business Associate Agreements with their telephone companies. The reason was structural, not political: the telephone carrier did not *possess* the PHI. It transmitted it. The carrier

was a conduit — a passive channel through which information moved — and its access to that information was transient, incidental, and non-persistent. The carrier could not read the content of a call; it merely routed the signal.

This intuition was formalized in the HIPAA Omnibus Final Rule, published in the Federal Register on January 25, 2013, which codified the **conduit exception** to the business associate definition. [6] The Omnibus Rule clarified that the conduit exception is limited to *transmission-only* services for PHI, including any *temporary* storage of PHI incident to such transmission. The critical qualifier is the word "persistent": a conduit has only *transient* access to PHI. The moment a vendor's access to PHI becomes persistent — stored, processed, analyzed, or retained for any purpose beyond the immediate transmission — the conduit exception is unavailable, and the vendor is a business associate required to execute a BAA.

The Department of Health and Human Services (HHS) Office for Civil Rights has confirmed this distinction in authoritative guidance. In a published FAQ addressing cloud service providers, HHS stated unequivocally: "Generally, no. CSPs that provide cloud services to a covered entity or business associate that involve creating, receiving, or maintaining (e.g., to process and/or store) electronic protected health information (ePHI) meet the definition of a business associate, even if the CSP cannot view the ePHI because it is encrypted and the CSP does not have the decryption key." [7] The HHS guidance further specifies that "a CSP that maintains ePHI for the purpose of storing it will qualify as a business associate, and not a conduit, even if the CSP does not actually view the information, because the entity has more *persistent* access to the ePHI." [7]

The structural lesson is profound: the conduit exception is not about the *sensitivity* of the information, nor about the vendor's *intent* to access it. It is about the *nature of the vendor's relationship* to the data. Transient transmission = conduit. Persistent storage or processing = business associate.

B. Why AI Platforms Are Categorically Business Associates, Not Conduits

An AI platform is, by its fundamental architecture, the opposite of a conduit. When an attorney submits a prompt to an AI platform, the platform does not merely route the prompt from sender to recipient. It *processes* the prompt — ingesting it, analyzing it against its model weights, generating a response, and, in most consumer and enterprise configurations, *retaining* the prompt in session logs, training datasets, or

audit records. This is persistent access. It is the precise category of vendor relationship that HHS has confirmed falls outside the conduit exception.

The distinction maps cleanly onto the telecom analogy:

Characteristic	Telephone Carrier (Conduit)	AI Platform (Business Associate)
Access to content	None (routes signal only)	Full (processes and analyzes content)
Storage duration	Transient (milliseconds)	Persistent (session logs, retention periods)
Use of content	None	Generates outputs; may train models
BAA required?	No	Yes, if PHI is processed
Regulatory status	Conduit exception applies	Business associate definition applies

Table 8: Conduit vs. Business Associate — The Structural Distinction.

This analysis has immediate and practical consequences for attorneys who represent healthcare clients, handle medical malpractice litigation, or work on HIPAA enforcement matters. Any attorney who submits a prompt containing PHI — patient records, medical histories, insurance claim data, or any individually identifiable health information — to an AI platform that lacks an executed BAA has potentially committed a HIPAA violation on behalf of their client, regardless of whether the prompt is otherwise protected as work product.

C. The Evolution of HIPAA's Reach: From 1996 to the AI Era

HIPAA's evolution from a 1996 insurance portability statute to a comprehensive data privacy framework tracks almost precisely the evolution of the digital infrastructure through which health information flows. The original 1996 Act focused primarily on insurance continuity and contained relatively modest privacy provisions. The Privacy Rule (2003) and Security Rule (2005) established the substantive framework for PHI protection. The HITECH Act (2009) dramatically expanded enforcement authority and introduced breach notification requirements. The Omnibus Rule (2013) extended

direct liability to business associates and their subcontractors, and codified the conduit exception with the precision that now governs AI platform analysis.

Each of these evolutionary steps followed the same pattern: Congress and HHS identified a new category of entity that had persistent access to PHI and brought it within the regulatory framework. Clearinghouses, billing services, transcription companies, cloud storage providers — all were successively incorporated as the healthcare system's technological infrastructure expanded. AI platforms are the next category in this progression, and the regulatory framework already contains the tools to analyze them. The question is not whether HIPAA applies to AI-assisted legal work involving PHI; it does. The question is whether the attorney has taken the steps necessary to ensure that the AI platform they use has executed a BAA and operates as a compliant business associate.

D. Platform HIPAA Status: A Practitioner's Reference

The following table provides a streamlined HIPAA compliance reference for the platforms analyzed in this manuscript. It is designed for practitioner use rather than comprehensive regulatory analysis:

Platform	BAA Available	HIPAA-Ready Tier	Critical Limitation
Harvey AI	Yes	Enterprise (core platform)	Email Harvey, Web Browsing, and Knowledge Source subprocessors are explicitly excluded from BAA coverage
OpenAI (Enterprise/API)	Yes (application required)	ChatGPT Enterprise; ChatGPT for Healthcare; API (healthcare customers)	Consumer tiers (Free, Plus, Team) are not covered
Anthropic Claude	Yes (conditional on eligible API use)	Eligible Anthropic API endpoints (Messages API) with ZDR; Claude Code with enterprise API credentials	BAA does not cover Free, Pro, Max, or Team plans; Claude for Work/Enterprise chat interfaces are NOT ZDR-eligible; most beta features excluded
Microsoft 365 Copilot	Yes (within M365 BAA)	M365 Copilot; Copilot Studio; Copilot for Security	Must be configured within a HIPAA-ready M365 tenant
Google Gemini (Workspace)	Yes (Workspace BAA)	Gemini within Google Workspace	Standalone Gemini app is not covered; some features blocked for BAA customers
Manus AI (All Tiers)	**No**	**Not HIPAA compliant**	Terms of Use explicitly prohibit PHI; no BAA program; must not be used for healthcare matters involving PHI
Casetext CoCounsel	Not publicly documented	Not confirmed	No public BAA program confirmed
LexisNexis (Lexis+ AI)	Not confirmed for AI platform	Not confirmed	No public BAA program for Lexis+ AI confirmed

Table 8: AI Platform HIPAA Compliance Reference for Legal Practitioners.

The practical implication for attorneys is straightforward: before using any AI platform in connection with a healthcare matter, the attorney must (1) determine whether the matter involves PHI; (2) if so, confirm whether the platform has an executed BAA in place; and (3) if no BAA is available, either use a compliant platform or ensure that no PHI is included in any prompt submitted to the platform. The work product protection analysis of this manuscript applies independently of HIPAA — but HIPAA adds a separate and potentially more severe layer of regulatory exposure for attorneys who handle health information.

XI. PRACTITIONER CHECKLIST: PLATFORM DISCIPLINE FOR WORK PRODUCT PROTECTION

The following checklist distills the analysis of this paper into actionable practitioner guidance:

- **PRE-USE:** Identify the platform tier (consumer, enterprise, legal-purpose) before submitting any litigation strategy prompt.
- **PRE-USE:** Review and document the platform's current data retention and ZDR terms. Maintain a Platform Confidentiality Log.
- **PRE-USE:** Confirm existence and terms of any applicable enterprise DPA or customer agreement governing the platform relationship.
- **PRE-USE:** Score the platform on the AgFi Scorecard. Do not use platforms scoring below Silver for litigation strategy prompts.
- **AT USE:** Apply the TOLFPC Sandwich Defense: frame every strategy prompt with explicit anti-fabrication instructions at opening and close.
- **AT USE:** Generate prompts that disclose only what is necessary for the research objective. Minimize client identification in prompts whenever possible.
- **AT USE:** Maintain contemporaneous notes of AI research sessions, documenting date, platform, query categories (not full text), and output use.
- **POST-USE:** Apply the Chain-of-Verification protocol to all AI-generated legal propositions before reliance.
- **POST-USE:** Execute the Deletion Protocol: review and purge any AI-session records on platforms that do not auto-delete.

- **IF SUBPOENA:** Assert work product protection immediately in writing upon any discovery demand implicating AI platform records.
- **IF SUBPOENA:** File a FRCP 45(d)(3) motion to quash the subpoena. Assert opinion work product as near-absolute protection.
- **IF SUBPOENA:** Produce the Platform Confidentiality Log as evidence of maintained confidentiality expectation.
- **FIRM POLICY:** Implement a written AI Governance Policy designating approved platforms by tier and prohibiting consumer-tier AI for litigation strategy use.
- **HEALTHCARE MATTERS:** Before using any AI platform on a healthcare matter, confirm whether the matter involves PHI. If PHI is present, use only platforms with an executed BAA (Harvey AI, OpenAI Enterprise, Anthropic Claude via eligible API endpoints with ZDR, or Microsoft 365 Copilot in a HIPAA-configured tenant). Note that Claude for Work and Claude for Enterprise chat interfaces are not ZDR-eligible; BAA coverage requires use of eligible Anthropic API endpoints.
- **HEALTHCARE MATTERS:** Do not use Manus AI, Casetext CoCounsel, or Lexis+ AI for any prompt containing PHI until those platforms establish a public BAA program.
- **HEALTHCARE MATTERS:** Even on BAA-covered platforms, avoid features explicitly excluded from BAA coverage (e.g., Harvey AI's Email Harvey and Web Browsing subprocessors).

XII. COUNTERARGUMENTS AND SCHOLARLY RESPONSES

A manuscript that proposes a new doctrinal framework has an obligation to engage with the strongest objections to that framework. Four counterarguments warrant serious scholarly attention.

A. The "Mere Tool" Objection: AI Is Not an Agent

The strongest objection to the Heppner Protective Shield framework is that the *Kovel* analogy fails because an AI platform is not an agent in any legally cognizable sense. *Kovel* extended privilege to an accountant because the accountant was a human professional exercising judgment in service of the attorney's legal representation. An AI platform, the objection runs, is a tool — no different from a word processor or a legal

research database. Attorneys do not assert work product protection over their Westlaw queries; why should AI prompts be different?

This objection is analytically serious but ultimately unpersuasive for three reasons. First, the *Kovel* doctrine does not require the agent to be human. The Second Circuit's reasoning in *Kovel* focused on function, not form: the accountant was protected because the accountant's work was necessary to the attorney's legal representation and was directed by the attorney in service of that representation. An AI platform that generates legal strategy analysis at the attorney's direction, under the attorney's supervision, and in anticipation of litigation satisfies the same functional criteria. Second, the analogy to Westlaw queries fails because Westlaw queries are not expressions of attorney cognition — they are search terms. A litigation strategy prompt, by contrast, contains the attorney's legal theory, anticipated adversarial arguments, factual analysis, and strategic reasoning. The prompt is the attorney's mental impressions rendered into text. The *Hickman* categories protect mental impressions regardless of the medium in which they are expressed. Third, the *Warner* court explicitly rejected the "mere tool" characterization, holding that AI-assisted drafting materials are protected work product and that "ChatGPT (and other generative AI programs) are tools, not persons" — a holding that, while framed in the negative, functionally treats AI as an instrument of attorney cognition rather than an independent actor, which is the precise framing that the Heppner Protective Shield framework adopts.

B. The Waiver Objection: Submission to the Platform Is Disclosure

A second objection holds that an attorney who submits a prompt to an AI platform has voluntarily disclosed the prompt to a third party — the platform — and that this disclosure waives work product protection. Under this view, the attorney's submission of litigation strategy to ChatGPT is no different from mailing a strategy memorandum to a stranger.

This objection misapplies the waiver doctrine. Waiver of work product protection requires disclosure "in a manner inconsistent with maintaining secrecy from adversaries." *In re Steinhardt Partners, L.P.*, 9 F.3d 230, 235 (2d Cir. 1993). Disclosure to a litigation support vendor — a document management firm, a discovery service provider, a jury consultant — does not waive work product protection because it does not make the material available to the adversary. The AI platform, used in the same manner as a litigation support vendor, falls within the same analytical framework. The objection has more force with respect to consumer-tier platforms, where the absence

of a confidentiality agreement creates arguable space for a waiver challenge — which is precisely why HPS Prong 4 requires confidentiality architecture as a formal element of the protection analysis.

C. The Novelty Objection: Courts Have Not Adopted This Framework

A third objection — the most methodologically precise — holds that the HPS Test is aspirational, not descriptive, and that presenting it as a doctrinal framework overstates the current state of the law. No court has adopted the HPS Test. The AgFi scoring system has no judicial pedigree. The TOLFPC protocols are the author's own constructs.

This objection is correct as a matter of positive law, and this manuscript acknowledges it explicitly. The HPS Test is a proposed framework, not a binding standard. The AgFi scoring system is an analytical construct, not a judicially adopted metric. The manuscript's claim is not that courts *have* adopted this framework, but that the analytical factors the framework organizes are the same factors courts are *already* applying — implicitly and without a unified vocabulary. The contribution of the manuscript is to supply that vocabulary, not to describe a legal reality that already exists in fully articulated form. The appropriate scholarly response to the novelty objection is not to abandon the framework but to be precise about its status — which this manuscript endeavors to be throughout.

D. The Platform Independence Objection: Commercial Bias in Platform Ratings

A fourth objection holds that the manuscript's favorable ratings for specific AI platforms (Harvey AI, Casetext CoCounsel, Westlaw AI, Lexis+ AI) may reflect commercial relationships rather than independent analysis. The objection is that a framework that recommends specific commercial products is advocacy, not scholarship.

This objection is addressed by the methodology underlying the platform ratings. Every rating in Table 2 is based on publicly available technical and contractual documentation: the platform's published privacy policy, its enterprise DPA terms, its ZDR commitments, and its BAA availability. No platform has been rated favorably on the basis of any commercial relationship. Attorneys should independently verify current platform terms before reliance, as terms of service change. The framework's value is in the analytical criteria it establishes — ZDR status, DPA architecture, BAA

availability, AgFi scoring — not in the specific ratings assigned to specific platforms at the time of writing.

XIII. LIMITATIONS AND SCOPE

This manuscript advances a doctrinal framework for a rapidly evolving area of law. Intellectual honesty requires explicit acknowledgment of its limitations.

Jurisdictional Scope. The analysis in this manuscript is grounded in federal law — specifically, FRCP 26(b)(3), the *Hickman* line of Supreme Court authority, and the circuit-level cases cited throughout. State court work product doctrine varies significantly across jurisdictions. Attorneys practicing in state court proceedings should conduct independent analysis of applicable state rules before relying on the HPS framework.

Temporal Scope. The AI work product jurisprudence is developing rapidly. *Heppner* and *Warner* were decided in February 2026. By the time this manuscript reaches its audience, additional decisions may have refined, extended, or contradicted the analytical framework advanced here. The platform data in Table 2 reflects publicly available information as of March 2026 and will require updating as platform terms evolve.

Empirical Limitations of AgFi Scoring. The AgFi scoring system is an analytical construct based on the author's assessment of publicly available platform documentation and observed AI performance characteristics. It has not been validated through controlled empirical testing. The scores assigned in Table 4A are illustrative, not definitive, and reasonable practitioners may score specific platforms differently based on their own experience. The framework's value is in the analytical dimensions it identifies, not in the specific scores it assigns.

The HPS Test's Unresolved Prong 4 Questions. Prong 4 of the HPS Test — the confidentiality architecture requirement — raises questions that this manuscript does not fully resolve. What level of enterprise DPA protection is sufficient to satisfy Prong 4? Does a DPA that prohibits model training but permits retention for audit purposes satisfy the prong? Does a platform's SOC 2 Type II certification, standing alone, satisfy the prong? These questions will require judicial resolution, and this manuscript does not purport to answer them definitively. It identifies the analytical framework; courts will supply the specific answers.

HIPAA Analysis Scope. The HIPAA analysis in Section X is limited to the business associate / conduit distinction and the BAA requirement. It does not address the full scope of HIPAA compliance obligations for attorneys, including the HIPAA Security Rule's technical safeguard requirements, the Breach Notification Rule, or the intersection of HIPAA with state privacy laws. Attorneys handling healthcare matters should consult qualified HIPAA counsel in addition to the analysis provided here.

No Legal Advice. This manuscript is a work of legal scholarship. Nothing in it constitutes legal advice. Attorneys should apply the frameworks described here to their specific facts and circumstances, in consultation with applicable bar ethics opinions and, where appropriate, ethics counsel.

XIV. CONCLUSION: THE PROTECTION OF THE LEGAL MIND

The attorney who submits a litigation strategy prompt to an AI research platform — regardless of whether that prompt contains client confidences — has generated opinion work product under *Hickman v. Taylor* and FRCP 26(b)(3). This is not a novel proposition. It is the application of settled doctrine to a new medium — an application that the existing case law strongly supports, even if courts have not yet spoken with uniformity. Courts that have recognized this — *Tremblay*, *Concord Music* (N.D. Cal. 2025), *Warner* — have applied existing doctrine correctly to new facts. Courts that have denied protection have done so not because the doctrine is inapplicable, but because the attorneys before them failed to satisfy the structural conditions that the doctrine requires. The Heppner Protective Shield Test is the map of those conditions.

The protection is real, it is robust, and it is available in every federal circuit. What varies is not the existence of the protection but the practical burden of its defense. That burden is entirely eliminated by zero data retention, substantially reduced by enterprise-grade data processing agreements, and elevated — sometimes severely — by consumer-tier platform use with default retention. **Exposure is not loss of protection.** It is the transformation of automatic protection into contested protection. Sophisticated practitioners do not accept that transformation when it can be avoided.

The profession has not yet fully grasped this distinction. Many practitioners assume that AI prompts containing no client confidences require no privilege analysis. They are wrong — and the first contested subpoena directed to a consumer AI platform seeking litigation strategy prompts will make the consequences of that assumption concrete and costly. The *Heppner* decision illustrates the consequences of

unstructured AI use. The *Warner* decision illustrates the protection available when structure is present. This manuscript proposes the architecture for achieving that structure consistently.

This manuscript argues that the Heppner Protective Shield is the most doctrinally sound architecture currently available for structuring AI-assisted legal work. The HPS Test has not yet been formally adopted by any court. What the existing case law demonstrates, however, is that courts are already applying the same five analytical factors — attorney direction, anticipation of litigation, cognitive externalization, confidentiality architecture, and non-waiver — when evaluating AI-generated materials. The HPS Test proposes to give those factors a coherent, named structure. An attorney who builds an AI-integrated litigation practice around these five factors is aligning their conduct with the analytical framework that the emerging case law implicitly applies. The framework is offered as a proposed doctrinal structure that courts and bar authorities should adopt as the AI work product jurisprudence matures.

For healthcare attorneys, the analysis extends further. The same platform-selection decision that determines work product protection also determines HIPAA compliance. The conduit exception — the principle that explains why physicians never signed BAAs with AT&T — strongly supports the conclusion that AI platforms are business associates, not conduits, and that attorneys who handle health information should treat the HPS framework as a predicate HIPAA compliance consideration. The convergence of work product doctrine, professional responsibility, and HIPAA within a single analytical framework is this manuscript's central contribution to the literature on AI-integrated legal practice.

Prompts, when constructed and deployed under the Heppner Protective Shield, are not external inputs. They are extensions of the attorney's legal cognition — the attorney's mental impressions, legal theories, and litigation strategy rendered into text through the medium of artificial intelligence. *Hickman* protected those mental impressions in 1947. FRCP 26(b)(3)(B) protected them in 1970. The Heppner Protective Shield protects them in the age of AI.

The legal mind has always been the attorney's most protected asset. The Heppner Protective Shield ensures it remains so.

REFERENCES

[1] *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. 2023) (Castel, J.) (imposing sanctions on attorneys who submitted AI-generated brief containing six fictitious case citations without verification). Available at Justia. [2] *United States v. Heppner*, 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026) (Rakoff, J.) — The decision is available on Westlaw at the foregoing citation. For secondary commentary, see also: Loose AI Prompts Sink Ships: How Heppner Shook the Legal Community, N.Y. State Bar Ass'n (2026) [3] *Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03223-AMO, 2024 WL 3748003 (N.D. Cal. Aug. 8, 2024) (Martínez-Olguín, J.) (holding that AI prompts written by lawyers constitute opinion work product because they "were queries crafted by counsel and contain counsel's mental impressions and opinions about how to interrogate [an AI tool]"); *Concord Music Group, Inc. v. Anthropic PBC*, No. 24-cv-03811-EKL (SVK), 2025 WL 1482734 (N.D. Cal. May 23, 2025) (affirming same principle; case originally filed in M.D. Tenn. and transferred to N.D. Cal., where the May 2025 work product ruling was issued); *Concord Music Group, Inc. v. Anthropic PBC*, 2025 WL 2267950 (N.D. Cal. Aug. 8, 2025) (holding that AI prompts generated by non-legal corporate employees, as opposed to counsel, are discoverable — reinforcing the attorney/non-attorney distinction). For secondary commentary, see also: Barbara Tsao & Amy Heath, *A Closer Look: The Discoverability of Artificial Intelligence Prompts*, Inside Class Actions (Feb. 25, 2026), https://www.insideclassactions.com/2026/02/25/a-closer-look-the-discoverability-of-artificial-intelligence-prompts/. [4] *Warner v. Gilbarco, Inc.*, No. 2:24-cv-12333, 2026 WL 373043 (E.D. Mich. Feb. 10, 2026) (Patti, M.J.) (denying motion to compel discovery of pro se plaintiff's AI tool interactions; holding that AI-assisted drafting materials are protected work product; the court stated that "ChatGPT (and other generative AI programs) are tools, not persons"). For secondary commentary, see also: Proskauer Rose LLP, *Michigan Federal Court Protects AI-Assisted Litigation Work Product* (Mar. 2, 2026), https://www.proskauer.com/alert/michigan-federal-court-protects-ai-assisted-litigation-work-product. [5] ABA Comm. on Ethics & Prof'l Responsibility, Formal Op. 512, *Generative Artificial Intelligence Tools* (July 29, 2024) (addressing lawyers' ethical obligations under Model Rules 1.1 (competence), 1.4 (communication), 1.5 (fees), and 1.6 (confidentiality) when using AI tools). See also: *ABA Issues First Ethics Guidance on a Lawyer's Use of AI Tools*, Am. Bar Ass'n (July 29, 2024), https://www.americanbar.org/news/abanews/aba-news-archives/2024/07/aba-issues-first-ethics-guidance-ai-tools/. [6] Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act, 78 Fed. Reg. 5,566 (Jan. 25, 2013) (codified at 45 C.F.R. pts. 160, 164) (the "HIPAA

Omnibus Rule"). Available at https://www.federalregister.gov/documents/2013/01/25/2013-01073/modifications-to-the-hipaa-privacy-security-enforcement-and-breach-notification-rules-under-the. [7] U.S. Dep't of Health & Hum. Servs., Office for Civil Rights, *FAQ: Can a Cloud Service Provider Be Considered a "Conduit" Like the Postal Service, and Therefore Not a Business Associate?*, HHS.gov (last visited Mar. 2026), https://www.hhs.gov/hipaa/for-professionals/faq/2077/can-a-csp-be-considered-to-be-a-conduit-like-the-postal-service-and-therefore-not-a-business%20associate-that-must-comply-with-the-hipaa-rules/index.html ("Unlike a conduit, a CSP that maintains ePHI has persistent access to the data, both the encrypted and unencrypted data, and therefore must enter into a BAA with the covered entity."). [8] Steve Alder, *HIPAA Conduit Exception Rule and Transmission of PHI: 2026 Update*, The HIPAA Journal (Jan. 2, 2026), https://www.hipaajournal.com/hipaa-conduit-exception-rule/. [9] Harvey AI Corporation, *Platform Agreement* (last updated Jan. 9, 2026), https://www.harvey.ai/legal/platform-agreement ("Subprocessors will not train any AI models using Your Content or Customer Data. Subprocessors, except for cloud storage providers, will not retain or log content for human review."). *Note:* This ZDR commitment applies at the subprocessor level. Attorneys should independently verify current platform terms before reliance, as terms may be updated. [10] Anthropic, *Zero Data Retention (ZDR)*, Claude API Docs, https://platform.claude.com/docs/en/build-with-claude/zero-data-retention ("ZDR applies to the Claude Messages and Token Counting APIs" and "Claude for Work and Claude for Enterprise product interfaces are not ZDR-eligible"); Anthropic Privacy Center, *I have a zero data retention agreement with Anthropic. What products does it apply to?*, https://privacy.claude.com/en/articles/8956058 ("the only products to which zero data retention applies are eligible Anthropic APIs, and Anthropic products that use your Commercial organization API key"); Anthropic, *Business Associate Agreements (BAA) for Commercial Customers*, Anthropic Privacy Center (last updated Mar. 20, 2026), https://privacy.claude.com/en/articles/8114513-business-associate-agreements-baa-for-commercial-customers. *Note:* BAA coverage depends on qualifying services and configurations; attorneys should confirm that their specific use case and service tier are covered before relying on Anthropic's BAA for HIPAA compliance. [11] Paubox, *Is Manus AI HIPAA Compliant? (2025 Update)* (Dec. 12, 2025), https://www.paubox.com/blog/is-manus-ai-hipaa-compliant-2025-update. *Note:* As of the date of this manuscript, Manus AI's Terms of Use explicitly prohibit submission of PHI. Attorneys should not use Manus AI for matters involving PHI without first

confirming the existence of a signed BAA and confirming that PHI submission is permitted under the applicable service tier.