Bayla Law

AI and Confidentiality: The Confidentiality Continuum, a Practical Guide for Lawyers

By Christian Andreu-von Euw and Jayson L. Cohen

Introduction

This article focuses on the *confidentiality* of AI in the legal world, not its *reliability*.[1] The guidance on reliability is voluminous, consistent, and uncontroversial: AI hallucinates, and lawyers must verify any information it produces.

By contrast, only a few court decisions have addressed AI's confidentiality. They provide conflicting signals about court confidence in AI's ability to protect confidential information, but they also suggest a lack of technical understanding about how AI tools work among judges and lawyers. Despite this uncertain landscape, attorneys have duties of competence and confidentiality that they must fulfill when using AI or advising clients about its use. This article seeks to bring some clarity that will help lawyers understand and apply existing ethical and court guidance when they use AI and advise clients on AI risk mitigation and options.

The article has four parts. Part I presents the ethical rules governing the duties of confidentiality and competence that apply to AI, along with bar opinions interpreting those rules for internet services and AI. Part II examines the few court decisions that have considered AI's confidentiality, including the controversial opinion in *United States v. Heppner*. We review the split in authority on whether work product protection survives AI use and identify the tendency of courts to sort AI technology into "open" systems (deemed unsafe) and "closed" systems (deemed safe). To start moving beyond this open/closed dichotomy, Part III offers a high-level description of the technology underlying AI. Here, we distinguish the different behaviors by AI model providers, such as data storage and use, that affect confidentiality risk. Part IV then lays out the continuum of confidentiality choices available to lawyers and clients who want to use AI: from the "open" consumer product tier, to "closed" business and enterprise product tiers offered by major AI model providers, to even more protective product and service offerings.

I. Ethical Rules and Their Application to AI

Lawyers have duties to protect client confidences that apply to AI use. Model Rule 1.6 of the Rules of Professional Conduct addresses an attorney's obligation to keep client information

[1] The term "artificial intelligence" applies to many technologies including spam filtering, speech recognition, and aspects of traditional search engines. Here, it refers to software based on large language models such as those offered by vendors like OpenAI and Anthropic, often referred to as Generative AI.

confidential and applies to all technology use by lawyers, including AI.[2] Rule 1.6(a) prohibits an attorney from disclosing client information without informed consent, among other exceptions. All states have adopted an identical or similar rule. Rule 1.6(c) further requires "reasonable efforts to prevent the inadvertent or unauthorized disclosure of, or unauthorized access to, information relating to the representation of a client." Most states have adopted or are in the process of adopting a form of Model Rule 1.6(c).

These Rule 1.6 confidentiality obligations work hand-in-hand with Model Rule 1.1, which requires basic attorney competence. It includes a directive to "keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology."[3]

AI is not the first technology that entrusts client data to third-party providers, and past bar opinions have addressed such technologies under these rules. The use of email elicited bar opinions that generally find it is safe if the attorney makes reasonable efforts to protect client confidences. For example, ABA formal opinion 477R reaffirmed that *unencrypted* email can be an acceptable method of lawyer-client communication but re-emphasized that lawyers must understand the risks and tradeoffs of every form of online communication.[4] It concluded broadly that "lawyers must, on a case-by-case basis, constantly analyze how they communicate electronically about client matters, applying the factors [from comment 18 to Model rule 1.6] to determine what effort is reasonable":[5]

- the sensitivity of the information;
- the likelihood of disclosure if additional safeguards are not employed;
- the cost of employing additional safeguards;
- the difficulty of implementing the safeguards; and
- the extent to which the safeguards adversely affect the lawyer's ability to represent clients (e.g., by making a device or important piece of software excessively difficult to use).[6]

State bar associations have issued similar opinions allowing the use of third-party services like cloud software and online data storage as long as attorneys exercise care.[7]

---

[2] American Bar Association, *Model Rules of Professional Conduct* (2013).
[3] *Id.*, Comment 8 to Model Rule 1.1.
[4] ABA Standing Committee on Ethics and Professional Responsibility, Formal Opinion 477R: *Securing Communication of Protected Client Information*, at 3-5 (May 11, 2017).
[5] *Id.* at 5.
[6] *Id.* at 4.
[7] *See, e.g.*, Ill. State Bar Ass'n Prof'l Conduct Advisory Op. 16-06 (2016) ("A lawyer's use of an outside provider for cloud-based services is not, in and of itself, a violation of Rule 1.6, provided that the lawyer employs, supervises and oversees the outside provider."); Wash. State Bar Ass'n Advisory Op. 2215 (2012) ("a lawyer using online data storage must not only perform initial due diligence when selecting a provider and entering into an agreement, but must also monitor and regularly review the security measures of the provider."); N.C. State Bar 2011 Formal Op. 6 ("[A] lawyer may contract with a vendor of software as a service provided the lawyer uses reasonable care to safeguard confidential client information"); Pa. Formal Op. 2011-200 ("An attorney may ethically allow client confidential material to be stored in 'the cloud' provided the attorney takes reasonable care to assure [confidentiality and security]."); *see also* ABA Formal Op. 08-451 (2008) (foundational opinion allowing outsourced storage of client information).

Recent bar guidance that is specific to AI has built on this foundation. In 2023, the California State Bar issued guidance on applying the reasonableness standard to the use of AI.[8] It directed lawyers to review terms of use and consult an IT professional before inputting client information into an AI tool.[9] California is also revising the comments to its conduct rules, reiterating that AI does not alter a lawyer's duties of competence and confidentiality, which require the exercise of professional judgment and the protection of client information.[10] These duties also encompass the need to obtain informed client consent before using AI. ABA Formal Opinion 512, issued in 2024, advises that lawyers need informed client consent before inputting client information into an AI tool.[11] The New York City Bar Association similarly requires informed client consent.[12]

This ethical guidance steers in one direction: the better attorneys understand AI technology, the better they can satisfy their ethical obligations and serve their clients.

II. Confidentiality in the Courts

Most Courts Have Protected AI Use for Litigation as Work Product

Court guidance on confidentiality issues raised by AI has been less clear than bar guidance. Commentators have focused on *United States v. Heppner*, an early decision that allowed discovery of AI use, and denied claims of work product and privilege.[13] It led to countless law firm articles warning of AI's dangers, many of which gave the mistaken impression that all AI use is inherently risky or that general purpose commercial AI tools offer insufficient protections.

Commentators have paid less attention to *Warner v. Gilbarco, Inc.*, another decision that issued on the same day as *Heppner* and reached the opposite conclusion.[14] Since then, most, but not all, court decisions have protected the inputs and outputs of AI use from discovery in litigation. Below, we present an overview of the cases that have considered this issue, providing a starting point for lawyers trying to make decisions about their own AI use, advise a client about the client's AI use, or advocate for their preferred AI practices in Court.

Heppner held that a defendant's interactions with a "publicly available AI platform" were neither attorney-client privileged nor entitled to work product protection under the Federal Rules of

[8] The State Bar of California, Standing Committee on Professional Responsibility and Conduct, *Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law*, at 2-3 (Nov. 16, 2023).
[9] *Id.* at 2; *see also id.* (regarding Rule 1.1, Duty of Competence: "Before using generative AI, a lawyer should understand to a reasonable degree how the technology works, its limitations, and the applicable terms of use and other policies governing the use and exploitation of client data by the product.").
[10] The State Bar of California, *Proposed Amendments to the Rules of Professional Conduct Related to Artificial Intelligence*, https://www.calbar.ca.gov/public/public-meetings-comment/public-comment/public-comment-archives/2026-public-comment/proposed-amendments-rules-professional-conduct-related-artificial-intelligence.
[11] ABA Standing Committee on Ethics and Professional Responsibility, Formal Opinion 512: *Generative Artificial Intelligence Tools*, at 6-7 (July 29, 2024).
[12] New York City Bar Association, Committee on Professional Ethics, Formal Opinion 2024-5: *Gen AI in the Practice of Law*, at 3 (Aug. 2024).
[13] 820 F. Supp. 3d 292 (S.D.N.Y. 2026) (memorializing Feb. 10 oral opinion).
[14] 820 F. Supp. 3d 629 (E.D. Mich. 2026).

Criminal Procedure.[15] It reached that conclusion after considering the consumer-tier terms of service for that AI tool and the types of disclosure that were possible in that situation. The court was particularly concerned about the AI model provider's ability to "train" its models on Heppner's user data, an important issue we discuss below. *Warner* reached the opposite conclusion. It held that the work-product doctrine under the Federal Rules of Civil Procedure did protect a pro se litigant's use of a similar consumer AI tool, finding that the use of AI did not constitute a disclosure to an adversary, which would be required to waive work product protection.[16]

Three subsequent decisions found *Warner*'s AI-friendly analysis more persuasive than *Heppner*'s analysis. *Morgan v. V2X Inc.* applied *Warner* to hold that work-product protection in a civil case applied to use of AI tools.[17] It analogized use of AI to the expectation of privacy in everyday use of internet-based email and messaging through third-party providers like Google.[18] In *Assini v. Hayward*, a New York court expressly followed *Warner* and *Morgan* and quashed a subpoena seeking a litigant's ChatGPT logs, holding that AI use neither surrendered confidentiality nor waived work-product protection.[19] And in *Tate Group Automotive, LLC v. Legacy Automotive Capital, LLC*, a Texas court similarly followed *Warner* and *Morgan* and found that a non-lawyer's "chats" with ChatGPT were protected work product, but also ordered disclosure of all discovery materials shared with ChatGPT.[20]

Two courts, however, have followed *Heppner*. *Fry v. Fry* found that uploading to Claude under Anthropic's consumer terms of service destroyed the secrecy needed to justify sealing the uploaded documents.[21] And *Shealy v. Seaside Investments, LLC* denied work product protection because a civil litigant's counsel was not involved in the AI use; rather, the romantic partner of the litigant used ChatGPT on his behalf without counsel's knowledge.[22]

In contrast to this post-*Heppner* split of legal authority, two decisions that issued before *Heppner* and *Warner* protected an attorney's use of AI during litigation. *Tremblay v. OpenAI, Inc.* and *Concord Music Group, Inc. v. Anthropic* both blocked production of lawyer prompts and their corresponding AI responses.[23, 24]

Protective Orders Offer Guidance

Court decisions that consider prospective protective order limitations on AI use provide insight into how courts understand AI technology. They also provide guidance on where courts may draw lines in the future to separate protected use from waivers of work product and privilege. These court decisions appear to divide AI applications into two classes based on concerns about

[15] 820 F. Supp. 3d at 296-99.
[16] 820 F. Supp. 3d at 636-37.
[17] No. 25–cv–01991, 2026 WL 864223, at *3-5 (D. Col. Mar. 30, 2026).
[18] *Id.* at *4.
[19] No. 607683/2024, --- N.Y.S.3d ---, 2026 WL 1677232, at *3-4 (Sup. Ct. Nassau Cty. June 4, 2026).
[20] Order, No. 25-BC11B-0020, slip op. at *1-4 (Tex. Bus. Ct. June 3, 2026) (Dorfman, J.).
[21] *Fry v. Fry*, No. 26-cv-3469-KSM, 2026 WL 2531956, at *3 n.4 (E.D. Pa. Aug. 27, 2026).
[22] *Shealy v. Seaside Investments, LLC*, Order, No. 2684CV00799-BLS2, slip op. at *2-6 (Mass. Sup. Ct. Suffolk Cty. June 16, 2026) (Squires-Lee, J.).
[23] No. 23-cv-03223, 2024 WL 3748003, at *2-3 (N.D. Cal. Aug. 8, 2024).
[24] No. 24-cv-03811, 2025 WL 1482734, at *1-3 (N.D. Cal. May 23, 2025).

disclosing confidential information: "open," "no-cost," or "consumer" AI tools are deemed to raise a real risk of training on and disclosure of confidential information, while "closed" or "enterprise" AI tools sufficiently eliminate these risks. The *Morgan* case, discussed above, issued a protective order forbidding "mainstream low-to-no-cost AI" tools whose terms allow "(1) storing or using inputs to train or improve its model; [or] (2) disclosing inputs to any third party except where such disclosure is essential to facilitating delivery of the service," while permitting the parties to use more protective "Enterprise-tier AI" instead.[25] *Jeffries v. Harcros Chemicals, Inc.* similarly issued a protective order forbidding the use of "open AI tools" on discovery materials but permitting "closed AI tools."[26] And *Fry* denied protection of materials disclosed under Anthropic's "non-enterprise" *consumer* terms of service.[27]

The chart below provides a summary of the case law discussed above.

Confidentiality of AI Use: Case Law

Case	User	Context	Use Protected?	Case(s) Followed
Tremblay v. OpenAI, Inc.,	Lawyer	Discovery	**Yes**	—
Concord Music Grp., Inc. v. Anthropic PBC,	Lawyer	Discovery	**Yes**	—
Warner v. Gilbarco, Inc.,	Non-lawyer	Discovery	**Yes**	—
United States v. Heppner,	Non-lawyer	Discovery	**No**	—
Jeffries v. Harcros Chems. Inc.,	Lawyer	Protective Order	**Yes ("closed" tools)**	—
Morgan v. V2X, Inc.,	Non-lawyer	P.O. & Discovery	**Yes (TOS dependent)**	Warner
Tate Grp. Auto., LLC v. Legacy Auto. Cap., LLC,	Non-lawyer	Discovery	**Yes**	Warner/Morgan
Assini v. Hayward,	Non-lawyer	Discovery	**Yes**	Warner/Morgan
Shealy v. Seaside Invs., LLC,	Non-lawyer	Discovery	**No**	Heppner
Fry v. Fry,	Non-lawyer	Sealing	**No**	Heppner

Reality is more nuanced than "open" and "closed" AI. The many choices about data storage, data use, and terms of service and use present a continuum of AI confidentiality possibilities. These may influence the advice an attorney provides to a client. We present this continuum of choices below. But, first, so attorneys can better understand these choices, we take a step back to explain how the large language models at the core of AI work and how they are accessed.

III. Technology: AI Basics and Confidentiality

Several court opinions discussed above, and many commentators, share the belief that AI poses a unique and inherent confidentiality risk because AI models are trained on user data. The primer

[25] 2026 WL 864223, at *7.
[26] No. 25-2569, 2026 WL 820218, at *4 (D. Kan. Mar. 25, 2026).
[27] *Fry*, No. 26-cv-3469-KSM, 2026 WL 2531956, at n.4.

below on AI technology shows there is no such inherent risk because AI training and subsequent AI use are separate and unrelated processes.

AI Models, their Training, and Inference

A large language model is, at its core, an enormous set of mathematical equations that receives text and predicts the text most likely to follow.[28] A simple version is found in modern email tools that auto-predict short phrases (like "Attached please find") from just a few characters (like "Att"). More sophisticated models work on the same principle, even if vastly more complex and capable.

AI models use data in two basic phases: (1) training, when the model learns to recognize patterns; and (2) inference, when the model generates outputs "inferred" from the inputs.[29] In training, the AI model adjusts its settings with the goal of predicting patterns in the training data. Any data used in training can influence the settings. This creates a risk that input training data may, directly or indirectly, be reflected in later outputs.[30]

In inference, end-users, such as lawyers and their clients, interface with the model. Users input a prompt (also known as a query or request), which the model processes to produce an output. Importantly, no training occurs during inference when using the AI models offered by major model providers.[31] The model's settings remain fixed.

While user inputs and outputs during inference do not affect the model, logs of these inputs and outputs, kept by the model provider, can be, and sometimes are, used to train new models or new versions of the same models. This training, if it occurs, creates some risk that user information could potentially be disclosed to a subsequent user of that model. Because of this, unsurprisingly, legal authorities largely agree that confidential information should not be input into an AI offering that may train its model(s) on user session data.

As explained in Part IV below, training on user logs can be disabled in almost all AI products, and it is generally disabled by default in AI products offered to businesses.[32]

---

[28] *See* OpenAI, *Better Language Models and Their Implications*, OpenAI (Feb. 14, 2019), https://openai.com/research/better-language-models; *see also* Anthropic, *Tracing the thoughts of a large language model* (Mar. 27, 2025), https://www.anthropic.com/research/tracing-thoughts-language-model.

[29] Nvidia, *What's the Difference Between Deep Learning Training and Inference?* (updated 2025), https://blogs.nvidia.com/blog/difference-deep-learning-training-inference-ai/.

[30] *See, e.g.*, Nicholas Carlini et al., *Quantifying Memorization Across Neural Language Models*, International Conference on Learning Representations (Mar. 6, 2023), https://arxiv.org/abs/2202.07646.

[31] *See, e.g.*, Anthropic, *Choosing a Claude model and effort level in Claude Code*, https://claude.com/blog/claude-model-and-effort-level-in-claude-code (last visited July 20, 2026) ("The weights of each model are set during training, and by the time you're sending requests they're read-only. Nothing in your prompt, your CLAUDE.md, or your context changes them. (If you've run into the word inference, that's all it means: using the model after training is done, with the weights fixed.)").

[32] AI algorithms that update a model from user inputs during inference exist but are rare. They are, however, outside the scope of this paper because the major model providers represent that their commercial models do not train on user inputs. *See* OpenAI, *Business Data Privacy, Security, and Compliance*, https://openai.com/business-data/ (last visited July 25, 2026) ("we do not use . . . inputs or outputs [] for training or improving our models"); Anthropic, *Is My Data Used for Model Training?*, Anthropic Privacy Center, https://privacy.claude.com/en/articles/7996868-is-

AI Tool Design

Users rarely interact with AI models directly. Instead, software tools mediate the exchange, translating a user's requests into one or more inference steps and assembling the results into a coherent response.

The simplest such tool is the AI chatbot, which captures the user's prompt, sends it to a model on the model provider's servers, and displays the response. The user experiences a somewhat normal conversation. But this simple design hides complexity. The model remembers nothing from one request to the next.[33] To create the appearance of an ongoing conversation and increase the usefulness of its answers, the chatbot logs and resends the accumulated inputs and outputs from the entire session with each new request. This is an example of "context," a term of art used to describe the working information given to an AI model along with a request.[34] It helps the model tailor its answers to the specific task and situation instead of producing a general response.

Many chatbots add other context, such as details about the user, with the aim of providing more tailored results. This context comes from account settings, user-saved instructions and preferences, user profiles, and "memory" files generated by the AI tool for the chat.[35]

Context management can give the user the impression that the model is learning from its interactions, as if the user were training the model by chatting. But no training is occurring. The model is never modified, and the information being collected is personal to the user's chat or account. It does not affect how the model behaves with other users.

Specialized software tools, such as legal AI tools, add more context. They combine the user's requests with information the user may never see, such as pre-configured instructions designed to steer the model toward a particular goal. A legal research tool, for example, might instruct the model to rely only on authority from its own case law database, to cite holdings rather than dicta, and to flag any proposition it cannot tie to a source. Generally, the more specialized the software, the more it layers additional context onto the basic exchange of query and result.

---

my-data-used-for-model-training (last visited July 25, 2026) ("we will not use your inputs or outputs . . . to train our models"); Google, *How Gemini for Google Cloud Uses Your Data, Google Cloud Documentation*, https://docs.cloud.google.com/gemini/docs/discover/data-governance (last updated July 17, 2026) ("Gemini doesn't use your prompts or its responses as data to train its models").

[33] *See* Claude Docs, *Using the Messages API*, https://platform.claude.com/docs/en/build-with-claude/working-with-messages (last visited July 25, 2026) ("The Messages API is stateless, which means that you always send the full conversational history to the API.").

[34] *See, e.g.*, Anthropic, *Effective Context Engineering for AI Agents* (Sept. 29, 2025), https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents.

[35] *See* OpenAI, *Memory FAQ*, https://help.openai.com/en/articles/8590148-memory-faq (saved memories, in addition to chat history, are "part of the context ChatGPT uses to generate a response") (last visited July 26, 2026); OpenAI, *Dreaming: Better Memory for a More Helpful ChatGPT* (June 4, 2026), https://openai.com/index/chatgpt-memory-dreaming/.

IV. AI Product Tiers – the Confidentiality Continuum

Major Model Product Tiers & Their Features and Confidentiality Safeguards

With this basic understanding of how AI works, the confidentiality of an AI model provider's product boils down to several questions:

- What user data does the AI model provider store and for how long?
- What can the model provider do with user data?
- What are the model provider's contractual terms and policies governing that storage and use?

The major AI model providers tend to group their products and services into four tiers, each of which provides different answers to these questions: (1) consumer products, (2) business and enterprise products, (3) API access, and (4) zero data retention. These tiers offer an increasing level of confidentiality protections and control, and each of Anthropic, OpenAI, and Google offers a product or service that falls within each tier.

Consumer Products

At the least protective end of the confidentiality continuum are ***consumer products***. These typically save inputs and outputs on cloud servers and are often accompanied by terms of use and user settings that allow model providers to use user data, including inputs and outputs, to train future models.[36] By default, training on user chat logs is often enabled for consumer accounts, but it can typically be disabled in the user settings.[37]

Users can further mitigate use of session data by deleting individual chats or using a "temporary" or "incognito" chat mode. When a user takes these precautions, the model provider retains no

[36] *See* OpenAI, *How Your Data Is Used to Improve Model Performance*, https://openai.com/policies/how-your-data-is-used-to-improve-model-performance/ (Mar. 13, 2026) (ChatGPT "improves by further training on the conversations people have with it, unless you opt out"); Anthropic, *Updates to Consumer Terms and Privacy Policy* (Aug. 28, 2025), https://www.anthropic.com/news/updates-to-our-consumer-terms ("We will train new models using data from Free, Pro, and Max accounts" when the model-training setting is on); Google, Gemini Apps Privacy Hub, *Gemini Apps Privacy Notice*, https://support.google.com/gemini/answer/13594961 (last visited July 26, 2026) ("Google uses [your] data . . . to: . . . improve our services . . . [.] These uses extend to the generative AI models . . . *. Please don't enter confidential information that you wouldn't want a reviewer to see or Google to use to improve our services* . . . ."; before June 29, 2026, this page said: "Google uses your activity to . . . improve its services (including training generative AI models)").

[37] *See* OpenAI, *How your data is used to improve model performance*, https://help.openai.com/en/articles/5722486 (last visited July 26, 2026) ("When you use our services for individuals such as ChatGPT and Codex, we may use your content to train our models. You can opt out of training through our privacy portal . . . ."); Anthropic, *Consumer Terms of Service* (Oct. 8, 2025), https://www.anthropic.com/legal/ consumer-terms ("We may use [inputs and outputs] to provide, maintain, and improve the Services and to develop other products and services, including training our models, unless you opt out of training through your account settings."); Google, Gemini Apps Privacy Hub, *Gemini Apps Privacy Notice*, https://support.google.com/gemini/answer/13594961 (last visited July 26, 2026) ("If the Keep Activity setting is off and you don't submit feedback, Google does not use your future chats to improve our AI models.").

permanent record of the chat history. Temporary or incognito mode is never used for training, and deleting a chat prevents its future use for training although it cannot unwind prior training.

Although training is disabled upon a deletion request, the deletion of the information is not immediate. Deleted or incognito-mode data is retained on the model provider's backend — for safety, compliance, and operational purposes — for a brief period ranging from 72 hours to 30 days, depending on the provider.[38] But providers carve out exceptions to these retention rules for safety-flagged content and legal holds.[39] These types of deletion delays and exceptions are also common in other cloud services.

Retained user inputs, outputs, and logs are also subject to disclosure terms in consumer terms of service or privacy policies that are more permissive than in the commercial products discussed below. For example, Anthropic's consumer privacy policy provides that "[w]e may share personal data [including user inputs and outputs] with government authorities, law enforcement, or other third parties where, based on the information available to us, we have a good-faith belief that disclosure is reasonably necessary to (i) comply with applicable law, regulation or legal process," among other reasons.[40] The *Heppner* court cited a previous version of this policy to justify its finding that AI inputs were not confidential.[41]

Business and Enterprise Products

A step up from consumer products are ***paid business accounts*** (with names like "Team" or "Business"). This is the first tier with characteristics that seem to fit the "closed" paradigm adopted by some courts. Products in this tier typically disable training on user data by default.[42] They also have terms of service with more robust privacy protections.[43] For example, Anthropic's business tier terms are more restrictive than the consumer tier disclosure terms

---

[38] *See* OpenAI, *Chat and File Retention Policies in ChatGPT*, https://help.openai.com/en/articles/8983778-chat-and-file-retention-policies-in-chatgpt (last visited July 22, 2026) (deleted chats and Temporary Chats removed within 30 days unless de-identification or legal exceptions apply); Anthropic, *Using Incognito Chats*, https://support.claude.com/en/articles/12260368-using-incognito-chats (last visited July 22, 2026) (incognito chats retained for 30 days by default, longer with custom enterprise retention settings); Google, Gemini Apps Privacy Hub, *Gemini Apps Privacy Notice*, https://support.google.com/gemini/answer/13594961 (last visited July 26, 2026) (Temporary Chats and Activity-off conversations retained up to 72 hours; conversations selected for human review may be retained up to three years even after user deletion).

[39] For example, certain OpenAI consumer chat data, referred to as output log data, are currently subject to expanded legal hold retention. *See* Stipulation and Order to Terminate OpenAI's Ongoing Obligations Under the Preservation Order at ECF 33, *The New York Times Co. v. Microsoft Corp.*, No. 23-cv-11195 (S.D.N.Y. Oct. 9, 2025), ECF 922 (terminating Preservation Order issued in umbrella multidistrict litigation, No. 25-md-3143, and imposing more limited preservation obligation); Order Denying Objection, No. 25-md-3143 (S.D.N.Y. June 26, 2025), ECF 271 (affirming Preservation Order, No. 25-md-3143 (S.D.N.Y. May 13, 2025), ECF 33).

[40] Anthropic, *Privacy Policy* (July 8, 2026), https://www.anthropic.com/legal/privacy.

[41] *Heppner*, 820 F. Supp. 3d at 296 (discussing the Anthropic consumer privacy policy effective February 19, 2025, https://www.anthropic.com/legal/archive/a2eecf43-807a-4a53-89dd-04c44c351138).

[42] *See* OpenAI, *How your data is used to improve model performance*, https://help.openai.com/en/articles/5722486 ("By default, we do not train on any inputs or outputs from our products for business users, including ChatGPT Team, ChatGPT Enterprise, and the API"); *Generative AI in Google Workspace Privacy Hub* (May 26, 2026), https://knowledge.workspace.google.com/admin/generative-ai/generative-ai-in-google-workspace-privacy-hub ("Workspace does not use customer data for training models without customer's prior permission or instruction.") Anthropic, *Commercial Terms of Service* (June 17, 2025), https://www.anthropic.com/legal/commercial-terms ("Anthropic may not train models on Customer Content from Services.").

[43] *See id.* §§ C–E; *OpenAI Services Agreement* (Jan. 1, 2026), https://openai.com/policies/services-agreement/ § 7.

criticized in *Heppner*. They require Anthropic to be subject to a legal process seeking disclosure of the data before it can disclose, and if permitted, it must provide pre-disclosure notice of that process to the user.[44]

As with consumer products, AI model providers store business account chat inputs and outputs by default. For practical reasons, this storage is desirable as it allows users to access their own work from any device and preserve it for future use. But business account users concerned about the security of this storage can delete chats or use incognito chats, subject again to retention periods.

A significant feature that typically distinguishes business products from consumer products is the availability of applications that can store history and chat files on a user's computer as opposed to the model provider's cloud location. Such local storage is often called a local *sandbox*. Applications that focus on users who develop software, such as Claude Code, OpenAI's Codex CLI, and Google's Gemini CLI, are notable examples of locally sandboxed environments.[45] These products are more customizable than those with standard chat interfaces and, despite being marketed as "code" tools, they suit a wide range of attorney use cases. Another example of a local storage tool is the "Local Projects" option included in the Anthropic business-focused Cowork application.[46] (Lawyers should exercise care with Cowork local projects because Anthropic is actively developing its Cowork offerings: currently, local storage must be expressly enabled.) Products that offer local storage still send data to the model providers' servers for inference, but the data is only retained there for a period comparable to the applicable API-retention window (discussed below).

In a higher tier are the ***enterprise accounts***. An enterprise product builds on the base confidentiality protections provided by a business product. Per above, training is disabled by default, history is stored on the provider's servers unless sandboxed, deletions and temporary chats are subject to a retention period, and disclosure to third parties requires legal process and, if permitted, pre-disclosure notice.

Beyond this, they offer customizable features and control that may enhance protections for confidential information, such as advanced user management, audit logging and compliance for Application Programming Interfaces, stronger security, expanded compliance options (e.g.,

---

[44] Anthropic's commercial terms define customer inputs and responsive outputs as Confidential Information, stating: "Recipient may disclose Discloser's Confidential Information to the extent it is required by law, or court or administrative order, and will, except where expressly prohibited, notify Discloser of the required disclosure promptly and fully cooperate with Discloser's efforts to prevent or narrow the scope of disclosure." Anthropic, *Commercial Terms of Service* § E.3, https://www.anthropic.com/legal/commercial-terms; *see also OpenAI Services Agreement*, https://openai.com/policies/services-agreement/ §§ 7.3, 17 (defining Confidential Information to include customer inputs and received outputs and providing "Recipient may disclose Confidential Information to the extent required by law, if Recipient uses reasonable efforts to notify Discloser, to the extent permitted, prior to doing so.").

[45] Claude Code stores history and certain other uploaded and AI-generated files locally to the folder ~/.claude/projects/. https://code.claude.com/docs/en/data-usage. OpenAI Codex CLI stores locally to ~/.codex/sessions/. https://github.com/openai/codex/discussions/8339. Gemini CLI stores locally to ~/.gemini/tmp/<project_hash>/chats/. https://github.com/google-gemini/gemini-cli/blob/main/docs/cli/session-management.md.

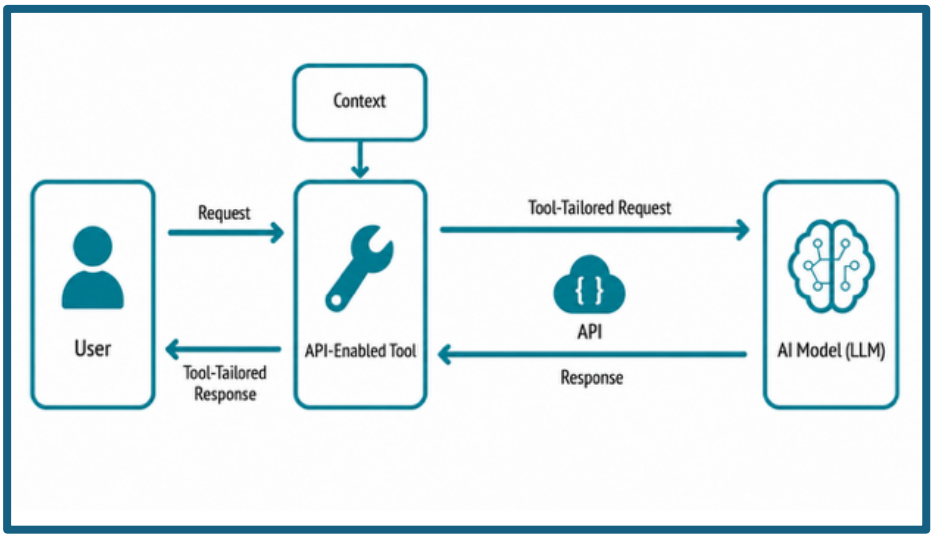
[46] https://support.claude.com/en/articles/14116274-organize-your-tasks-with-projects-in-claude-cowork (Cowork "[p]rojects are desktop-only and stored locally. There's no cloud sync for project data at this time.") (last visited July 30, 2026).

HIPAA and bespoke data processing and service level agreements), and higher capacity operation.[47] And larger enterprises can negotiate custom terms of service that modify the protections in other ways.

API Access

The major AI model providers all offer another account type that provides access to LLMs through Application Programming Interfaces (APIs). This allows third-party tool vendors (with their own operating systems and data storage locations) to use the highly capable models created by the major model providers. Although most attorneys do not use API access directly, many use it indirectly when they use Westlaw CoCounsel, Lexis Protégé, Harvey, Legora, and other commercial legal AI tools. These tools, like most third-party AI tools, use API access to models from the major model providers as the backbone of their software. (These specific legal AI tools and others will be discussed in a future article.)

The graphic below illustrates the architecture for an API-enabled tool. A user starts a user session in an API-enabled tool and makes a request. The tool tailors the request for the AI model according to (1) the user request, (2) the tool's specific functionality, and (3) the request's "context." The model's response may also be tailored by the tool before presented to the user.



API-enabled tools often allow their users to access and select between different models on a task-by-task basis. This means the tool user may need to be aware of the confidentiality rules and terms of service governing each model and governing the tool vendor's relationship with each model provider.

API access raises a slightly different set of issues around confidential information because it involves little-to-no data storage by the model provider. The model processes an input on receipt

[47] *See* OpenAI, *Enterprise privacy at OpenAI* (Jan. 8, 2026), https://openai.com/enterprise-privacy (SOC 2, SSO/SAML, SCIM, audit logs, data residency, HIPAA BAA); https://www.anthropic.com/enterprise (SSO, SCIM, RBAC, audit logs, Compliance API, custom retention; HIPAA BAA); https://knowledge.workspace.google.com/admin/generative-ai/generative-ai-in-google-workspace-privacy-hub (admin controls, Cloud DPA, HIPAA BAA coverage for Gemini for Workspace) (last visited July 15, 2026).

and returns an output, and the model provider immediately marks all data for deletion. The inputs and outputs are not used for training by default and are stored only briefly by the model provider according to its API retention policy (absent a zero data retention arrangement, discussed below).[48]

Users must also be aware that the vendor of the API-enabled tool has its own retention policies and may retain user data longer than the AI model provider, allowing that data to remain available to the user. Again, this retention is both practical and necessary as it allows users to access their own data on different devices and preserve the data for future use. Thus, attorneys using AI offerings that rely on API access need to be concerned with the retention and confidentiality policies of both the API-enabled tool vendor and the model provider.

Zero Data Retention

An even more protective way to use an AI model is through a zero data retention (ZDR) agreement on an enterprise or API plan. In this tier, the model provider eliminates the retention period altogether and contractually agrees not to store or review any input or output data. A ZDR option is not offered off-the-shelf by any AI model provider, likely because a model provider cannot assure itself that its models are not being used for dangerous or illegal activities if there are no records. Even enterprise products do not typically include a standard ZDR option. Instead, model providers require potential customers to negotiate ZDR agreements individually.[49]

ZDR may also be available for tool vendors who purchase an API access plan for a model. Large tool vendors, like legal AI tool vendors, often enter into such ZDR agreements with their model providers. The tool users can take advantage of these ZDR agreements, preventing storage of their data by the model provider. But as suggested above, these tool vendors typically store their users' data as well, to make it perpetually available. So, a ZDR agreement by itself does not eliminate the concerns associated with online storage of user data.

With this understanding of the products from the major AI model providers, lawyers can make informed choices for themselves and their clients. The following chart summarizes these choices.

---

[48] *See, e.g.*, Anthropic, *How long do you store my organization's data?*, https://privacy.claude.com/en/articles/7996866-how-long-do-you-store-my-organization-s-data (API inputs and outputs retained for 30 days by default); OpenAI, *Enterprise privacy at OpenAI,* https://openai.com/enterprise-privacy (same); Gemini API, *Vertex AI Privacy*, https://ai.google.dev/gemini-api/docs/interactions-overview (Interactions API has configurable retention of 1 to 55 days for paid tier; 1 day for free tier) (all last visited July 26, 2026).

[49] *See, e.g.*, Anthropic, https://privacy.claude.com/en/articles/8956058-i-have-a-zero-data-retention-agreement-with-anthropic-what-products-does-it-apply-to (June 9, 2026) (discussing Zero Data Retention agreements); OpenAI, *Enterprise Privacy*, https://openai.com/enterprise-privacy (ZDR available for API customers with qualifying use).

Confidentiality of AI Use: Choices (Major AI Model Product Tiers)

Tier	Protection	Long-term storage	Key traits
Consumer Claude Free/Pro · ChatGPT Free/Plus · Gemini (Free/Pro)	**Moderate (if training disabled)**	Cloud	Training often on by default (can disable); more permissive disclosure terms (some vendor discretion)
Business / Team Claude Team · ChatGPT Team/Business · Gemini for Workspace (Business)	**High**	Cloud	Training off by default; strong terms (legal process + notice before disclosure)
Enterprise Claude Enterprise · ChatGPT Enterprise · Gemini Enterprise	**High**	Cloud	Business-level confidentiality, plus admin controls, audit logging, HIPAA & other compliance, higher capacity
Business / Team / Enterprise — local storage Claude Code· OpenAI Codex CLI · Gemini CLI Claude CoWork – Local Project	**Very High**	Local	Same as business/enterprise, but only short-term cloud storage
API access	**Very High**	Local	Business-like confidentiality, but only short-term cloud storage

Business-tier products offer strong protection of client confidential information for most attorney use cases. They should be sufficient to satisfy ethical obligations and stave off allegations of disclosure. Although some commentators have expressed a belief that lawyers need "enterprise"-grade products, that is usually not the case. Some lawyers need HIPPA certifications or another enterprise-only feature. But for many, the marginal advantages of an enterprise-tier product may not justify the cost under Model Rule 1.6(c)'s "reasonable efforts" standard. This is especially true for solo law practices and boutique law firms that may be unable to obtain or to afford a license for an enterprise account. Rule 1.6(c) allows an attorney to consider, among other factors, "the cost of employing additional safeguards" and the "extent to which the safeguards adversely affect the lawyer's ability to represent clients."[50] And, at bottom, business tier products provide a high degree of protection for client confidential information.

AI Options that are More Protective than Major Model Provider Products

The products above, including ZDR, provide *confidentiality by contract*, in which the user is relying on the model provider's promise of confidentiality. Two other types of AI products go further, providing *confidentiality by design* in which the user has technological assurances that data will remain confidential. *Secure computing* offers mathematical security through encryption, and *local hosting* offers physical security.

Secure Computing

A secure computing system is engineered so the AI model provider is incapable of accessing user data during the process of inference or otherwise. It employs specialized hardware that keeps input and output data encrypted, making it impossible for the model provider or anyone else with physical access to the server to access the data. The cryptographic methods used also

[50] Comment 18 to Model Rule 1.6.

allow third parties to independently verify the data security. In this way, mathematically verifiable privacy replaces contractually promised privacy.

Meta and Apple offer secure computing. Meta's *Private Processing* architecture routes AI features on WhatsApp to a secure computing server hosting an AI model whose data Meta cannot access.[51] Apple's *Private Cloud Compute* routes Apple Intelligence requests to a hardware-isolated secure computing server hosting an AI model that Apple cannot inspect.[52] The upcoming version of MacOS will rely on similar cryptographically-protected computing using Google Gemini models.[53] This highly confidential architecture could be attractive to lawyers.

The Tinfoil.sh tool is another example of secure computing. It provides API access to AI models that are hosted and execute on secure cryptographic hardware, whose security has been verified by third parties. Tinfoil.sh offers the use of models whose settings are public (and can therefore be used by anyone without the cost of a business or higher tier account to use proprietary models). These public models come from many sources, including Meta, OpenAI, and Google.[54] That the models are "public" does not mean that Tinfoil.sh tool compromises the security of the user data input into those models for inference.

Until recently, public models have lagged one or two generations behind the leading proprietary models.[55] But newer public models released in China appear to be close to achieving parity with the most advanced proprietary American models.[56]

Local Hosting

The second option at this highest *confidential by design* tier is a firm-hosted AI model on firm-controlled network resources. The inference phase happens inside the firm's environment; inputs, outputs, and metadata never leave the firm's control.

This locally-hosted option has gone from impossible to something that most IT professionals and technically proficient lawyers can accomplish. Tools like Ollama, LM Studio, and MLX-LM let

---

[51] Private Processing, Meta's architecture for WhatsApp AI features, processes user data in a Confidential Virtual Machine (a type of Trusted Execution Environment). It uses anonymous credentials, OHTTP relays, and stateless processing such that messages are not retained. *See* Engineering at Meta, *Building Private Processing for AI tools on WhatsApp* (Apr. 29, 2025), https://engineering.fb.com/2025/04/29/security/whatsapp-private-processing-ai-tools/.
[52] Apple's Private Cloud Compute is the cloud component of Apple Intelligence; it processes user requests inside dedicated, hardened server nodes built on Apple silicon, with cryptographic attestation and anonymizing relays designed so that Apple itself cannot access user data. *See* Apple Security Research, Blog, *Private Cloud Compute*: *A new frontier for AI privacy in the cloud* (June 10, 2024), https://security.apple.com/blog/private-cloud-compute.
[53] https://blog.google/company-news/inside-google/company-announcements/joint-statement-google-apple/.
[54] Tinfoil supports a range of public models including Meta's Llama 3.3 70B, OpenAI's GPT-OSS 120B, Google's Gemma 4 31B, and DeepSeek V4 Pro. Tinfoil offers private AI inference using these public models inside secure hardware enclaves using NVIDIA confidential-computing-enabled GPUs and AMD SEV, with cryptographically verifiable guarantees that user data cannot be accessed by Tinfoil, the cloud provider, or any third party. Tinfoil, *Chat Models*, https://docs.tinfoil.sh/models/chat. Anthropic does not currently publish public versions of its models.
[55] *See* Jack Edwards and Luke Emberson, Epoch AI, *Open models lag state-of-the-art closed models by 4 Months* (May 29, 2026), https://epoch.ai/ data-insights/open-closed-eci-gap (open models trail by a single model generation); Stanford HAI, *2026 AI Index Report: Technical Performance*, https://hai.stanford.edu/ai-index/ 2026-ai-index-report/technical-performance (top closed model leads top open model by ~3.3% as of March 2026).
[56] Laurie Chen and Aditya Soni, https://www.reuters.com/world/china/a-new-inexpensive-chinese-ai-model-is-catching-up-with-anthropic-openai-their-2026-07-02/.

an attorney install and run a public model on a modern personal computer in minutes. For some tools, no command-line expertise is required for installation and use: for example, LM Studio has a graphical interface. A modern computer configured with ample amounts of RAM can comfortably run advanced models suitable for many tasks.

Of course, the use of local tools and models trades one risk for another, as the operator of the hardware must configure the models and ensure the security of its system. For less technologically capable users, it may be more secure to rely on a contract to protect information than to rely on one's technical skill.[57] Indeed, the outsourcing of security and its related risks are often seen as primary benefits of cloud-based vendor software.

Conclusion

Before choosing an AI tool, lawyers seeking to make reasonable efforts to protect their clients' confidential information must weigh the sensitivity of the information, the likelihood of disclosure, the cost of additional safeguards, and the extent to which potential safeguards could limit their ability to represent their clients. Consumer-tier products may be adequate for low-sensitivity tasks, and business-tier products are likely sufficient for most tasks that require confidentiality. But more protection may be needed in some highly specialized cases. For example, a lawyer working with data from a regulated industry may require an enterprise-tier or higher protection product. And a lawyer with representations that are likely to routinely trigger safety reviews by an AI model provider, such as representing clients that manufacture weapons or whose cases involve evidence of child sexual abuse, may need a locally hosted AI solution or another highly confidential environment.

In a future article, we will delve into the security and confidentiality issues associated with agentic AI and API access to AI models. We will also discuss how these issues apply to the legal AI tools available to lawyers such as Harvey and Legora, Westlaw Co-Counsel and Lexis Protégé, Claude Legal, and various alternatives. And we will address connectors and other plug-ins to software used by attorneys that expand not only the capabilities of that software but also the potential issues to consider when exercising reasonable efforts to protect client confidentiality.

Authors

Christian Andreu-von Euw

is a Partner at Bayla Law. His practice focuses on disputes involving artificial intelligence, computer security, trade secrets, patents, technology contracts, and other technology-focused issues.
He can be reached at christian@baylalaw.com

[57] Tools that simplify local model installation include Ollama (https://ollama.com), LM Studio (https://lmstudio.ai), MLX-LM for Apple Silicon (https://github.com/ml-explore/mlx), and llama.cpp (https://github.com/ggml-org/llama.cpp). Hardware capacity for local model inference scales with available memory (RAM on Mac, VRAM on PC GPUs): a useful rule of thumb is that a 7B-parameter quantized model requires approximately 4–8GB of memory, a 27B model approximately 16–24GB, and a 70B model approximately 40–48GB.

Jayson L. Cohen

is a Partner at Bayla Law. His litigation practice focuses on patents, trade secrets, artificial intelligence, and complex commercial disputes concerning high tech, telecom, life sciences, and chemistry. His space practice encompasses space and earth station licensing as well as rulemakings, coordination, and consultations involving satellite spectrum rights and use. He can be reached at jcohen@baylalaw.com.

Published August 6, 2026; Last updated September 15, 2026