

The Unguarded Mind

AI Chat Logs and a New Frontier for Discovery

CHRIS SCHWEGMANN

The author is the managing partner of Lynn Pinker Hurst & Schwegmann LLP, Dallas.

Sigmund Freud abandoned hypnosis because it produced the wrong kind of truth—revelatory in the moment but not integrated, not owned. Free association was different. It worked because it was voluntary: The patient chose, in a state of relaxed candor, to say what he or she actually thought.

When Benjamin Heppner, a financial services executive under federal investigation, opened the consumer version of Anthropic's Claude and began working through his legal exposure, he was not performing. He was thinking. He typed out the facts of his situation. He tested potential defenses. He probed his own vulnerabilities. He asked the kinds of questions—raw, specific, unvarnished—that people ask when no one is watching. He generated approximately 31 documents memorializing those exchanges. He later sent them to his lawyers at Quinn Emanuel, presumably as a kind of organized briefing book.

On February 17, 2026, Judge Jed S. Rakoff of the Southern District of New York handed every one of those documents to the federal government. In *United States v. Heppner*, the unedited candor in the chat logs became evidence. The ruling's doctrinal

take-away is simple. The litigation take-away is not: Artificial intelligence (AI) chat logs are the closest thing civil discovery has ever seen to a transcript of the unguarded mind.

Why AI Chats Induce Candor

But before we get to the playbook, it is worth pausing on the deeper question of why the AI chat produced the kind of candor that Heppner's situation called for—because understanding that psychology is what makes the discovery so valuable and what makes the failure to pursue it so costly.

In 1791, the English philosopher Jeremy Bentham designed a prison he called the panopticon. Its architecture was elegant in its cruelty: a circular structure with cells arranged around a central watchtower, designed so that any prisoner could be observed at any moment by a single guard—but, crucially, so that no prisoner could ever know whether he was being watched at any particular moment. The effect, Bentham argued, was that prisoners would eventually internalize the possibility of observation and police



their own behavior continuously. They would comply with the law not because they were being watched but because they could not be certain they were not.

The modern executive lives in a version of that world. Martin Lomasney, the Boston ward boss who ran machine politics in the West End for 50 years at the turn of the last century, understood the underlying principle long before e-discovery existed: Never write if you can speak; never speak if you can nod; never nod if you can wink. Decades of high-profile litigation taught sophisticated litigants that emails are permanent, texts are discoverable, Slack messages can be subpoenaed, and documents will be reviewed by people they have never met.

The AI chatbot broke that dynamic completely—and it did so by producing precisely the opposite of the panopticon effect. The interface was private and conversational. It felt, in every psychological sense, like thinking out loud alone. The user was not performing for an audience because, as far as he or she could tell, there was no audience. Bentham's guard tower was dark. The prisoner relaxed.

What came out when the prisoner relaxed is exactly what litigators spend careers trying to produce in a deposition room.

AI chat logs captured not just what people thought but what they were afraid of—the defenses they found unconvincing, the facts they could not reconcile, the theories they tested and abandoned, the questions they were not yet ready to ask their lawyers. It is a documentary form that seems to facilitate unguarded thought.

Consumer AI Is Collecting Potential Evidence

This is what makes *Heppner* so significant for litigators. The question is not really about privilege—it is about evidence. And the evidence that consumer AI platforms have been quietly accumulating is the closest thing to a transcript of the unguarded mind that courts have ever been asked to admit.

Start with what a litigator typically finds in discovery. The best emails—the ones that make careers and break cases—are the ones written before the writer knew they would matter. The message sent at 11:47 p.m. before the deal closed, not the carefully drafted follow-up the next morning. As institutional awareness of e-discovery has grown, the candid email has become rarer. People have adapted. They first moved sensitive emails to texts and then learned that texts are discoverable too. The panopticon

Illustration by Jon Krause

kept expanding, and sophisticated players kept finding the next dark corner.

Notes and journals occupy a different category—more candid, less filtered, occasionally devastating—but they suffer from a different limitation. They are retrospective. People writing in a journal are still performing, in a subtle way, for their future selves. They know they are creating a record, if only for their own eyes. The very act of keeping a journal implies some awareness of an audience, however imagined, however distant.

Artificial intelligence chat logs are the closest thing civil discovery has ever seen to a transcript of the unguarded mind.

Marcus Aurelius wrote his *Meditations* in Greek rather than Latin—his native tongue—possibly to further limit his readership. He was the most powerful person alive when he wrote it, and he wrote it entirely for himself: personal reminders, philosophical self-reproaches, the private thinking of a man working through problems he had not yet solved and principles he kept forgetting to follow. The Greek manuscript tradition calls it simply *Ta eis heauton*—“To Himself.” He was, as far as he knew, the only audience.

Two thousand years later, it is among the most widely read texts in Western history. What makes it so valuable to readers today is precisely what made Heppner’s AI documents valuable to a federal prosecutor: the pre-narrative quality, the unguarded self-examination, the questions people ask themselves *before* they have constructed an official position.

The deposition is litigation’s most direct attempt to access the unguarded mind, and it is the attempt that comes closest—which is precisely why good lawyers prepare their witnesses so thoroughly for it. The prepared deponent has a script, a vocabulary of non-answers, a trained instinct to pause before speaking. The five-second pause before answering a question is not a pause to remember. It is a pause to perform. The best deposition testimony is not candid testimony. It is controlled testimony, shaped by weeks of preparation, designed to give the questioner as little as possible while remaining technically truthful.

Ian McEwan’s 2025 novel *What We Can Know* imagines a pair of scholars in the year 2120—among the last historians alive after

cascading civilizational collapse—attempting to reconstruct a dinner party that took place in rural England in 2014. The fictional scholar Thomas Metcalfe works from the tools available to him: “ancient” emails, journals, search histories, and text messages that the guests left behind, their digital detritus preserved into the next century. From this, he reconstructs the arc of the evening, the tensions at the table, the conversations that preceded and followed the event. But he cannot quite close the gap between the record and the reality. The thoughts that were never typed, the fears that were never texted, the inner life that preceded the documentary trail—those remain just out of reach. Metcalfe, who is something of a romantic about his subjects, confesses at one point that he invents somewhat, filling the spaces between documents with inference, trying to honor what he calls his historical duty to vitality over the dead hand of academic neutrality.

Metcalfe also offers a warning to the people he studies—to us, that is—delivered with the urgency of someone who has spent a lifetime excavating the consequences of digital candor. His advice, addressed directly across the century: *If you want your secrets kept, whisper them into the ear of your dearest, most trusted friend. Do not trust the keyboard and screen.*

Benjamin Heppner did not take that advice. When Heppner opened Claude and began working through his legal exposure, he was not writing a document. He was doing something that felt, in every psychological respect, like thinking—the private, non-linear, genuinely exploratory process that precedes the moment when a person decides what he or she actually believes. And unlike everything in Metcalfe’s archive, the AI chat captured all of it—not the output of the thinking but the thinking *itself*, in real time, with no editorial distance between the fear and the text.

This is what makes the AI chat categorically different from any prior form of discoverable evidence. Judge Rakoff’s opinion does not describe in detail the contents of Heppner’s 31 documents. But the nature of the exercise Heppner undertook tells us a great deal about what they almost certainly contain.

First, they contain the questions—and the questions are often more revealing than the answers. When asking Claude, “Is it a problem that I knew about this transaction in advance?” or “What would the government need to prove to establish intent?,” a person is telling you, with unusual precision, exactly what he or she is worried about. The questions map the anxiety, identify the pressure points, and reveal, in the person’s own words and before any lawyer has shaped the narrative, what the person understands his or her exposure to be.

Second, the chat log contains the iterations—the moment when a person receives an answer he or she doesn’t like and pushes back, refines the question, or tries a different framing. In an ordinary document, you see the conclusion. In an AI chat, you see the reasoning process that produced it. You see the defenses that were considered and found wanting. You see the facts that

were introduced, then qualified, then walked back. You see the moments when the machine said something the user found unconvincing, and you see the user's response to that pushback. It is, in a very real sense, a real-time record of how someone thinks under pressure—which is what prosecutors and fact-finders most want to know and most rarely get.

Third, and perhaps most importantly, the AI chat contains what we might call pre-narrative material—the raw thinking that precedes the construction of the official story. By the time a client sits down with his or her lawyer for the first time, a narrative has already begun to form. It has been shaped by fear, by self-interest, by the instinct to present oneself favorably. The lawyer receives a version of events that has already been edited, consciously or not, by the client's survival instinct. The AI chat was created before that editing process began. It reflects the moment when the person was still asking, "What actually happened?" rather than "How do I explain what happened?" That difference—between the pre-narrative and the narrative—is the difference between the document that tells you what you need to know and the document that tells you what the party wants you to think.

The Implications of the *Heppner* Decision

Judge Rakoff's opinion captures this quality indirectly in its description of when the documents were created. Heppner generated the AI materials after receiving a grand jury subpoena and after it was clear he was the target of the investigation—but *before* his lawyers had directed any of the analytical work. That timing is not incidental. It is the signature of the most valuable class of AI discovery materials: the documents created in the period of high anxiety, before professional guidance has organized the client's thinking, when the person is most genuinely working through what he or she knows and what it means. The pre-representation danger zone, as it might be called, is exactly where the unguarded mind is most active—and exactly where the AI chat log is most likely to have captured it.

None of this is limited to criminal cases or to defendants. Consider the range of contexts in which the same dynamic plays out. An executive at a public company, upon learning of a government investigation, opens ChatGPT to work through what the investigators might find. A human resources director, after a discrimination complaint is filed, uses Claude to evaluate whether the termination decision creates legal exposure. A departing executive, anticipating a noncompete dispute, asks an AI platform to analyze the enforceability of the agreement and draft talking points for a conversation with a recruiter. A plaintiff in a breach of contract case uses Gemini to work through the damages theory before retaining counsel. In each case, the person created a detailed, candid, unguarded record of his or her

thinking—before the lawyers arrived, before the narrative was set, before the performance began.

The practical problem with AI discovery is not legal. Judge Rakoff settled the legal question with admirable efficiency—the documents are not privileged, they are not work product, and they are discoverable. The practical problem is that most litigators are not yet asking for them. The lawyers who move earliest, who serve preservation letters that explicitly name AI platforms before the opposing party has thought to clear their history, will get documents that later-moving counsel will never see.

The litigation hold letter needs updating most urgently. Most standard preservation letters enumerate categories of documents that have become conventional over the past two decades: emails, text messages, voicemails, internal communications, financial records. Almost none of them mention AI platforms. That omission is now a significant gap.

Going forward, any preservation letter in a complex litigation should include AI materials as a stand-alone, explicitly named category. The language does not need to be exotic. Something along these lines:

This letter constitutes notice of your obligation to preserve all communications with any artificial intelligence platform or tool, including ChatGPT, Claude, Gemini, Microsoft Copilot, Perplexity, and any other generative AI tool or service, relating to [the subject matter of this dispute], including chat logs, conversation histories, exported documents, and any materials generated through such interactions.

The specificity matters—a catch-all provision covering electronically stored information without naming AI platforms will not put an opposing party on adequate notice that the party's Claude history needs to be preserved, and it will not support a spoliation argument when that history turns out to have been deleted.

The document request is where most litigators will encounter AI materials for the first time—and the request needs to be drafted with enough specificity to reach the full range of what the opposing party may have generated. AI materials should not be buried in a general request for electronically stored information. They should appear as their own numbered request, or series of requests, identifying the distinct categories of material that AI interactions produce.

What Should Go into an AI Document Request

A well-drafted AI document request should reach at least four categories: First, chat logs and conversation histories with any AI platform relating to the subject matter of the dispute—including conversations that the party believes were deleted, to the

extent they remain recoverable. Second, documents exported or downloaded from AI platforms, including documents generated through AI interactions and subsequently saved, edited, or transmitted. Third, AI-generated drafts, analyses, memoranda, or strategy documents relating to the subject matter, regardless of whether those documents were subsequently shared with counsel. This third category is specifically designed to surface the *Heppner* pattern—the client who used AI to prepare analytical materials and then sent them to lawyers, creating the very documents that Judge Rakoff held to be outside the privilege. Fourth, any communications or documents reflecting instructions given to an AI platform, including prompts, follow-up questions, and refinements to prior queries, relating to the subject matter.

Before you can evaluate what you have—or challenge what you haven't received—you need to know what you're looking for. The interrogatories should establish, at minimum, which AI platforms the relevant witnesses used during the relevant period; whether those accounts were personal consumer accounts or enterprise accounts maintained by an employer; what data retention settings governed those accounts, including whether chat history was set to be preserved or automatically deleted; whether the witness used any AI platform in connection with the subject matter of the dispute, including any matter that eventually led to the present litigation; and the date on which the witness first became aware that litigation was reasonably anticipated.

Freud abandoned hypnosis because it produced the wrong kind of truth. A patient under hypnosis would say things that seemed revelatory in the moment but that the conscious mind had not yet processed, had not yet owned, and could not yet integrate into a usable understanding of the self. Free association was more valuable precisely because it was voluntary—because the patient chose, in a state of relaxed candor, to say what he or she actually thought. The technique worked not by bypassing consciousness but by creating conditions in which the conscious mind felt safe enough to speak without editing itself.

The AI chat interface created those conditions by accident. Nobody designed it to be a confessional. Nobody at Anthropic or OpenAI or Google sat down and said let's build a tool in which sophisticated people facing legal exposure will work through their most private fears and test their most vulnerable theories, and then generate a real-time transcript of their unguarded analytical process. The companies built a conversational tool that felt private and nonjudgmental and infinitely patient, and people used it the way people always use spaces that feel safe: They said what they actually thought.

But the deeper lesson of *Heppner* is not procedural. It is epistemological, and it is worth stating plainly before we return it to the shelf alongside Freud and Bentham and McEwan.

Litigation has always been, at its core, an epistemological project—an attempt to reconstruct what actually happened from the

incomplete and interested record that interested parties have left behind. The rules of evidence, the structure of examination and cross-examination, the entire apparatus of discovery—all of it is designed to close the gap between the narrative a party constructs after the fact and the truth that preceded it. The AI chat log closes that gap in a way that nothing in the prior history of litigation has managed to do.

McEwan's fictional Metcalfe, working from the rubble of our civilization a century hence, understood this intuitively. He understood that the letters and journals and emails and search histories his subjects left behind were remarkable—but that they still fell short of the inner life that produced them. He would

Going forward, any preservation letter in a complex litigation should include AI materials as a stand-alone, explicitly named category.

have found the AI chat log extraordinary, not because it replaced imagination, but because it made imagination unnecessary. The unguarded mind had spoken, in writing, before anyone thought to ask it to be quiet. McEwan's warning, delivered across a century by a fictional scholar in a dying civilization, was meant for us: If you want your secrets kept, whisper them to a trusted friend—or perhaps even a lawyer.

Benjamin Heppner did not take that advice. In a federal courthouse in Manhattan, the consequences are still unfolding. The litigators who understand what *Heppner* actually means are the ones who have understood something that most of their adversaries have not yet grasped: that the most important discovery in the next decade of litigation will come not only from email servers or text message archives. It will come from the quiet space where people went when they thought no one was listening, typed what they actually feared, and pressed Enter. ■

Chris Schwegmann can be reached at cschwegmann@lynnllp.com.