Minerva26

The Ruling Is In: GenAI Prompts Are Core Discoverable ESI

Hi, and welcome to this week's Case of the Week segment on the Meet and Confer Podcast. This is our final edition of the podcast for 2025. Looking forward to a fantastic 2026. I'm your host, Kelly Twigger, and this week we're doing something a little different. We look at case law because it's the only place we see how judges are actually applying the rules to real-world facts. And if we're not tracking those moves in real time, we're building strategy based on assumptions instead of reality.

So this week we're in the **OpenAI Copyright MDL** in the Southern District of New York — that's the consolidated litigation that includes the *New York Times Co. v. Microsoft & OpenAI* and other publisher cases.

At its core, the plaintiffs here say that OpenAI trained and runs ChatGPT on their copyrighted content without permission and is monetizing that through products like ChatGPT and Copilot.

Now instead of focusing on just one order, I want to walk through a **series of decision**s that all orbit around the same set of questions:

- How do you **preserve and produce chat logs at scale**?
- What do courts do when that collides with privacy promises and **data protection laws**?
- And, once the court orders a **20-million-chat sample**, what does it expect parties to actually do with the data?

There are **almost 20 discovery decisions** in this MDL since late 2024, including on mobile devices, social media, training data, and privilege. We're going to zoom in on the **ChatGPT conversation logs**—because this is widely viewed as the **first major ruling squarely treating prompts and outputs from a Generative AI system as core discoverable ESI**.

And as always, we're talking about this through the lens of what Minerva26 is for — helping you **map the data sources, understand the rules of the game, and decide what you're going to push or resist** when a court starts talking about logs, prompts, and outputs. That can be in the context of review with tools, or it can be in the context of actually using Generative AI tools where prompts are stored.

Part 1 – The court's first move: "Stop deleting chats"

Let's start at the beginning.

Early in the case, the news plaintiffs told the Court something that should make every in-house and outside counsel sit up:

> Open AI is **deleting ChatGPT output logs** — across consumer, enterprise, and API usage, including logs that users have asked to delete and logs subject to short retention windows.

The judge's reaction was basically:

> “I don't know yet exactly which of these logs are relevant. But I **do** know that I don't want them disappearing while we argue about it.”

So on **May 13, 2025**, Magistrate Judge Ona Wang entered a **Preservation Order** that essentially said to OpenAI:

- From **now on**, you must **preserve and segregate all output log data that would otherwise be deleted**.
- It doesn't matter if it would be deleted because of user requests, internal policies, or privacy regulations — you hold on to it until further order of the Court.

OpenAI came back with exactly the arguments you would expect from a company that's built its brand on privacy and safety:

- “We've made commitments to users about deletion.”
- “We have to comply with data protection regimes around the world.”
- “You're asking us to retain data explicitly that we promised to delete.”

The Court listened to those arguments and then **denied reconsideration**. District Judge Stein then later **upheld** the preservation order over OpenAI's objections.

So Phase One of the story is really very simple:

> Once your data is potentially relevant, and that includes prompts from generative-AI tools, courts will **override your normal deletion and privacy practices** in the name of preservation.

But I want to be clear about the *real* strategic takeaway here.

The lesson is **not** “wait until a judge thinks there's risk and then everything blows up.” By the time you're in front of a judge having this conversation, you may already have exposure for spoliation.

The more appropriate takeaway is:

As soon as litigation is reasonably anticipated, **part of your duty to preserve now includes understanding which tools your client uses to create prompts, where those prompts and outputs are logged, and the full scope of what may have been created that may be relevant**.

If you don't know that before you ever see a preservation order, you're already playing catch-up.

Part 2 – From "keep everything" to "what's a defensible sample?"

Once the judge has frozen the game with regard to preservation, the next question becomes:

> "All right, we've preserved everything, what are we *actually* going to use?"

And the point here in the MDL is that we're not actually going to review "all ChatGPT logs ever." That's not realistic.

And by mid-2025, everyone, plaintiffs and defendants, were all on the same page that the log universe from ChatGPT is enormous:

- OpenAI has **tens of billions of consumer output logs** and billions more from API use in the ordinary course.

So the fight shifts from **preservation** of those logs to a **sampling** of them.

And here's how that discussion played out:

- The plaintiffs asked for **120 million-log sample** — five million logs per month over two years — to test how the models behave in the wild.
- OpenAI pushed back hard on size and burden for that 120-million-log sample.
- And then OpenAI did something that became very important later:
    - It proposed **20 million logs** as the "reasonable" sample size.
    - It told the Court that the real cost is the time and compute to de-identify those logs.
    - And it insisted that using **OpenAI's own internal de-identification tool** was the right way to address privacy and personal information concerns.

Eventually, the plaintiffs agreed to that compromise:

> **20 million consumer logs,** pulled and de-identified by OpenAI.

So what does “**de-identify**” mean? I'm talking about how OpenAI describes this process:

> Running each log through an internal pipeline that strips or masks direct identifiers, which include names, email addresses, phone numbers, account IDs, etc., and some high-risk details so that you can no longer readily tie a conversation back to a specific person while still leaving the **prompts and outputs** intact enough to analyze how the system is used.

So by late summer in 2025, we have:

- A **fixed sample**: 20 million chats.
- A **defined process:** OpenAI will retrieve them from the archives, run them through its de-identification pipeline, and then the parties will work out the details of production under a protective order.

This is textbook proportionality in a big data case:

> Nobody is pretending that we can or should produce the entire universe. We negotiate a **statistically meaningful slice**, and then we build in a **privacy mechanism** on top of that.

Strategically, remember:

- That **20 million number is OpenAI's number**. They proposed it.
- Those **privacy assurances are OpenAI's assurances** about its own de-identification tool.

When you're in that negotiation for your clients, you have to assume that whatever you put out there as reasonable **may be the number that the court later locks you into**.

Part 3 – Narrowing preservation, clarifying why logs matter

Before we get to the big production fight, there are two bridging moves by the Court that really shaped the landscape here.

First, preserving less going forward

First, OpenAI and the plaintiffs eventually **stipulated to narrow** the forward-looking preservation burden.

The Court approved a deal that:

- Lets OpenAI **stop preserving new "would-be-deleted" logs** after a cut-off date;
- Required OpenAI to **keep the segregated block** of logs it had already preserved, and
- Required **ongoing preservation of logs tied to a set of publisher domains** — including the news plaintiffs, going forward.

So the preservation went from:

- "Everything that would have been deleted, going forward, indefinitely" to
- "Everything we've already segregated, plus logs related to these **specific domains**."

And that's a pretty pragmatic arc:

- From an emergency **freeze** at the beginning;
- Then a more **targeted, sustainable regime** once everybody understands the data better.

Second, logs matter for more than just "did you print my article"

Around the same time, District Judge Stein and Magistrate Judge Wang both underscore a key doctrinal point in related rulings:

> That the logs aren't just about "did this exact New York Times article appear in an output."

They are also about:

- **How people actually use the models in the wild**;
- Whether there are **substantial non-infringing uses**, which is OpenAI's own defense; and
- How outputs and usage patterns tie into the **fair use factor four**, which is market harm and damages.

Now that matters a tremendous amount for the sample that we're talking about:

- Because even logs where **no plaintiff content appears** can still be relevant to **OpenAI's defenses** and to the overall **damages and market-effect analysis**.
- You can't just say “If it doesn't contain the Times, it's irrelevant, so we can throw it away.” This is a key scope discussion. What is relevant here is relevant not only to the plaintiff's claims, but also to the defendant's claims of fair use as a defense.

So by the time we hit the production dispute, we've got this framework:

- A preserved chunk of logs from the earlier emergency regime narrowed going forward.
- A **20-million-log sample** that everyone has agreed to as the core slice for merits analysis.
- A **legal theory** that makes even “non-infringing” outputs relevant to both sides’ case theory.

Part 4 – The 20M-log fight: privacy vs. production (and the stay that never was)

The pivot: “We don’t want to produce all 20M anymore”

Now we get to the part everyone cares about: what happens when it's actually time to **produce** the 20 million logs?

Well, OpenAI has either finished or nearly finished de-identifying those logs with its own tool—the same tool that it touted to the Court as the right way to protect user privacy.

But then, OpenAI pivots.

Instead of introducing the entire 20 million sample, OpenAI tells plaintiffs it wants to:

- Run **additional search terms** against that set;
- Produce only the “**hits**”; and
- Treat the rest of the sample as effectively off-limits.

Plaintiffs then moved to compel **all** 20 million logs.

On **November 7, 2025**, Magistrate Judge Wang granted that motion and ordered OpenAI to **produce the full 20 million de-identified logs**, subject to the protective order, with room for the parties to negotiate additional privacy protections that don't stall discovery.

OpenAI then moved for **reconsideration**, which brings us to the **December 2, 2025** opinion — the one that really pulls together **relevance, privacy, and proportionality**.

Relevance: why the whole sample matters

On reconsideration to the magistrate judge, OpenAI leans into a frame that you're already hearing from clients:

- Plaintiffs supposedly admitted that "**99.9%**" of the sample is irrelevant;
- Only a tiny fraction of logs would ever have anything to do with the plaintiffs' works; and
- That makes production of the whole 20 million set wildly disproportionate.

The Court was not buying it.

In its December 2 order, Magistrate Judge Wang made two core points:

1. For **plaintiffs' claim**, the sample includes instances where their works appear—obvious relevance.
2. For **OpenAI's defenses**, the sample is central to:
    - **Fair use factor four** — impact on the market and damages; and
    - Whether the models are used for **substantial non-infringing purposes** across the user base.

And that second piece comes from **OpenAI's own theory of the case**—it has argued that the real-world usage shows substantial non-infringing use.

So the Court's message is, as you might expect:

> "You can't have it both ways, OpenAI. If you're going to say, 'look at how the models are used across the ecosystem' for your defense, you don't get to say that only the tiny slice where a plaintiff's article appears is relevant."

The 99.99% irrelevant framing gets rejected as unsupported and inconsistent with the record.

Privacy: one factor in proportionality, not a veto

OpenAI also pushed back hard on privacy, and this is where a lot of clients' instincts line up with what OpenAI said.

Their position, boiled down, is:

- The logs contain incredibly **sensitive user conversations**.
- We've marketed ourselves on respecting deletion and privacy.
- Producing a huge volume of chats, even de-identified, is a serious intrusion and should be narrowed to search hits.

Judge Wang doesn't really what we call caricature that. She calls the privacy concerns “sincere.” But she draws a very clear line:

- **Privacy is not a seventh factor in Rule 26(b)(1)**, the proportionality analysis. It's not even in the text.
- Instead, she says privacy is part of the **burden** and **risk** side of the proportionality analysis, and then asks whether the existing safeguards bring that burden down to a reasonable level.

Now, if you're a Minerva26 user or if you're interested in this article, please let me know. Retired Judge Francis wrote an article a couple of years ago doing a specific analysis on whether or not privacy should be included in the proportionality analysis. His conclusion was that it should not. And, here, we see that Judge Wang is including it as part of the burden analysis in that sixth factor of proportionality. So that's still an open issue. There's a lot of debate about that. It's an interesting consideration.

Judge Wang here explicitly said that **OpenAI had not shown** that its users’ privacy is not sufficiently protected by (1) the existing protective order and (2) OpenAI's exhaustive de-identification of the logs.

That's the key sentence to quote if you're going to use these cases or this particular decision from December 2 in any kind of presentation that you're giving about the discoverability of generative-AI, because privacy will not protect you from producing this information when you have those other steps in place.

Then Judge Wang walked through the protections that were already in place:

- First, **sampling** — We're not talking about all logs, we're talking about 20 million logs, which is a tiny fraction of what OpenAI keeps and a compromise from plaintiffs’ original ask of 120 million. Incidentally, that was also proposed by OpenAI.
- Second, **de-identification** — Every log in the sample goes through OpenAI's internal de-identification pipeline. OpenAI itself sold that tool as a more effective way to remove personal identifiers and private information than what the plaintiffs proposed initially.
- Third, **protective order and attorney's eyes only** — The logs are produced under a robust protective order with attorney's eyes only designations available. So only a narrow group of lawyers, experts, and vendors will ever see them.
- Finally, **more tailoring if needed** — The Court directed the parties to keep meeting and conferring about additional privacy measures that *do not further delay discovery*. And that further delay of discovery is what is critical.

Judge Wang also distinguished earlier social media chat decisions like—where limiting discovery to search term hits made sense because that's how relevance was framed—from this case, where the **usage patterns themselves** are part of the merits. And that's a very critical distinction, right? Where we're looking for specific terminology, we can use search terms, but

where we're looking for patterns and how a tool was actually used and how it recalled information, then those usage patterns themselves are part of the merits. So very different consideration here. And when you're looking at generative-AI content, you're going to have to consider that in your analysis.

The overall approach that Judge Wang took here is very consistent with other modern decisions on **social media and mobile devices**:

- Courts acknowledge these sources contain deeply personal content.
- But they treat privacy as a **limit on how** discovery proceeds—through scoped collection, filtering, de-identification, and protective orders—**not as a veto on whether** discovery happens at all.

Proportionality, credibility, and the December 9 stay fight

On proportionality, the Court stresses two pragmatic points.

- First, **scale** — 20 million logs is a **drop in the bucket** compared to OpenAI's retained logs — billions we talked about, tens of billions.
- The second, **marginal burden** — The heavy lift—retrieving, parsing, and de-identifying the logs has already been done by OpenAI or is already in progress. The incremental cost is actually producing what's been prepared.

Judge Wang also hinted at a credibility problem:

> If OpenAI never intended to produce the 20 million logs, why did it run its de-identification tool over all 20 million?
>
> If it did intend to produce them and then changed its mind, why should that change in strategy be rewarded?

So her bottom line was this:

- **Reconsideration was denied.**
- OpenAI was required to produce the full 20 million de-identified logs within seven days of completing de-identification, and to keep working with plaintiffs on privacy safeguards that don’t slow things down.

Next came the **December 9 order**, and this is where you want to emphasize OpenAI's **failure to get a stay**.

Here's the story:

- OpenAI has already lost on production on November 7, and on reconsideration on December 2.
- It had filed Rule 72(a) objections with District Judge Stein—but crucially **did not ask for a stay** at that time.
- At a December 4th hearing, OpenAI told Magistrate Judge Wang it expected to produce the logs "this week."
- Only **after** District Judge Stein later ordered more briefing on the objections did OpenAI send a letter to Magistrate Judge Wang asking her to effectively pause her own orders while Stein considered the issue.

On **December 9**, Judge Wang said no:

- Judge Wang noted that there was **no conflict** between her orders and Judge Stein's briefing order just because they both existed on the docket
- That the **Rule 72 objections do not excuse compliance** with a magistrate judge's discovery order unless a stay is actually granted.
- She cited authority for the basic proposition that you have to obey discovery orders while your objections are pending.
- She **denied** the stay request and reminded OpenAI that failure to comply could result in **Rule 37 sanctions**, including fee-shifting.

So as of right now:

The **production orders are in effect**.

- OpenAI's objections and appeals are pending in front of Judge Stein, but **there is, as of yet, no district court ruling** modifying, reversing, or staying the 20 million-log orders.
- That's important for you to be able to say cleanly in a client session. The briefing is still ongoing in front of Judge Stein, and we don't expect to have a ruling until Q1 2026.

Part 5 – What this means for strategy

So let's step back from all the weeds of those facts and talk about what you do with this, because our takeaways are really the point. And there are **three big takeaways** here to use with your team and your clients.

Takeaway 1 - Prompts are now squarely in the duty to preserve

If you wait until a judge says "ChatGPT prompts and outputs are relevant," you may already have months or years of **spoliation risk** exposure.

In this MDL:

- The **preservation** orders, which were—stop deleting output logs, segregate would-be-deleted data—came **before** the 20 million-log production orders.

If those preservation orders hadn't landed when they did, huge swaths of logs might simply not exist anymore.

So this is a point for both sides. If you're the plaintiff or the party that does not have this information, you need to be putting the other side on notice of its duty to preserve this information where it could be relevant. And, if necessary, go to the court early and get a ruling on these issues so that there is a preservation order in place.

It means that part of reasonable preservation now includes:

1. **Understanding the tools that are used to create prompts.**
    - ChatGPT, whether it's Free, Plus, Pro or Team accounts, ChatGPT Enterprise, Gemini, Claude, Copilot, in-house LLMs, and other embedded assistants are all fair game. Any outsource third-party tool that you might use that's generative-AI, also going to be discoverable.
    - Understand how users are accessing them, whether via browser, mobile app, Teams or Slack integrations, internal applications. How are they accessing them? Where is that information captured? Where are those logs available from?
2. **Understanding where prompts and outputs live and who can control them.**

    Individual users cannot flip a “legal hold” switch on the back end—but they can sometimes export their own data, and that's relevant to your strategy.

    - **ChatGPT consumer accounts** have a built-in “Export Data” feature under *Settings → Data Controls → Export*, which emails the user a ZIP containing their **entire chat history** and other account data.
    - That is effectively a **custodial export** at the user level. You *can* instruct individual custodians to run that export and preserve it as part of the litigation hold—just like you might with local PSTs historically.
    - But it is manual, time limited, i.e., the email link expires, and fragile at scale. It's not really a good substitute for proper system-level preservation, but it can work if that's what you need and you only have a few custodians.
3. **On the system side:**
    - For **OpenAI's own logs**, this MDL shows courts can compel the vendor to preserve and produce logs directly. That's non-custodial system-based discovery.

- In **Microsoft 365**, Teams/Copilot interactions live under the tenant-level and eDiscovery controls—legal holds are set by admins, not users. So non-user-based in Microsoft 365.
- In **Google Workspace**, Gemini data can be subject to Vault, but again, that's an administrative function that you'll need to investigate.
- API-based AI usage often creates logs in **your own databases and applications**, which means you have an internal system of record that is separate from the vendor.

4. **Custodian-based versus system-based discovery in this space**.

   We've talked before about moving away from strictly custodian-based discovery. This MDL is a very good example:

   - That **20 million log sample is not custodian-based**—it's system-based, drawn from the ChatGPT service across millions of users, because the question is, "How is this model used in the wild?" Which is a slightly pejorative statement, but one that comes up a lot because we're talking about how does everybody who uses ChatGPT, what do they use it for? And how is it drawing potentially on the system so that it can show how the potential copyrighted information that's at issue here was leveraged for that system's knowledge.
   - We still have **custodian-based** work in parallel—think Sam Altman's texts or DMs on his mobile devices. But for the logs, the unit of analysis is the **system**, not the named individuals.

5. **So, do you have to get to individual custodians for prompts?**

   - Sometimes, yes—user exports and screenshots can be relevant for a key witness.
   - But structurally, **generative-AI logs push you towards system-based discovery**: identify the platforms, understand their retention, and seek data at the service or tenant level, rather than reconstructing prompts custodian-by-custodian.

The duty to preserve now squarely includes **mapping those systems and deciding in advance how you'll preserve prompts and outputs**—vendor-side, tenant-side, or user-side—before you ever see a motion to compel.

That's exactly the kind of thing that we want to help you track in Minerva26. Not copies of the chats themselves, but the **map** of the tools, retention knobs, and which levers you can pull when a hold needs to go out.

Takeaway 2 - If you propose the "reasonable sample," you own it.

OpenAI is living in a world of its own making here:

- It proposed the 20 million chat sample as the right balance between relevance and burden.
- It promoted its own de-identification tool as the way to handle privacy.
- And then, once the logs were de-identified, it tried to claw back and produce only search term hits.

And the Court said:

> "Nope. You told us 20 million de-identified logs was reasonable. You did the work, now you're going to live with it."

So for your strategy:

- Be careful about the **numbers and the workflows** that you put into letters, joint statements, and ESI protocols.
- Assume that those numbers, as we've talked about many times here on Case of the Week, will become **anchors** in later proportionality rulings.
- Don't treat "we'll sample X" or "we'll de-identify in Y way" as throwaway talking points—assume the court will enforce them as commitments. We've seen that dozens of times here on what you put in writing is what the court will hold you to.

And remember the stay lesson from December 9:

- If you're going to fight a discovery order at the district judge level, **you have to think about a stay early**.
- Rule 72 objections alone do **not** suspend your obligation to comply; without a stay, you're expected to move forward while the objection is pending.

Takeaway 3 — Privacy is a dial, not an off switch.

The OpenAI rulings are a good snapshot of where courts are going on privacy and big data discovery and, in my view, where they already are:

- Privacy is **real** and gets explicit recognition.

- But it's treated as something you **engineer around**—with sampling, de-identification, protective orders, and AEO designations—not as an automatic escape hatch from discovery.

That tracks what we've seen for years with **mobile devices and social media**: judges don't just say “phones are private, so no discovery”; they say “phones are private, so we're going to do this carefully.”

It's the same thing here:

- De-identification reduces re-identification risk.
- Protective orders and AEO limit who sees what.
- And if you can show that the data set is central to the case, like these 20 million logs are to copyright, fair use, and damages, the court will **not** let privacy swallow relevance.

From a planning standpoint, this is why it helps to have a **strategic layer** sitting above your clients’ tools: something that keeps track of which systems are logging prompts, what privacy and retention settings exist, and what your preservation and production options are *before* you get to the motion to compel.

That's our Case of the Week-style look at the OpenAI MDL and what it teaches us about **preserving, producing, and arguing over generative-AI chat logs at scale**.

If your organization is using ChatGPT, Gemini, Claude, Copilot, or any other AI system that keeps logs, this is a preview of the kinds of orders you can expect to see—and the kinds of arguments on **relevance, privacy, proportionality, and stays** that are unlikely to carry the day.

If you're looking for a place to keep all of this case law strategy and matter intelligence in one place, that's exactly why we built Minerva26—to be the strategy layer that sits on top of your tools, not another system to feed. If you want to see how it could fit into your practice next year, reach out to us and we'll walk you through it using your real matters, not a demo script.

As we head into the holidays, I just want to say thank you for listening, for thinking hard with me about this stuff every week, and for the work that you do to make discovery smarter and more intentional for your clients. Here's to a little rest, a lot of joy, and coming back refreshed for whatever 2026 throws at us.

Thanks for joining me. I'm Kelly Twigger, and I'll see you next time on the Case of the Week.