

quinn emanuel trial lawyers | new york

295 5th Avenue, 9th Floor, New York, New York 10016 | TEL: (212) 849-7000 FAX: (212) 849-7100

April 23, 2026

MEMO ENDORSED

WRITER'S DIRECT DIAL
(212) 849-7164

WRITER'S EMAIL ADDRESS
andrewschapiro@quinnemanuel.com

VIA ECF AND EMAIL (FAILLA_NYSDCHAMBERS@NYSD.USCOURTS.GOV)

Honorable Katherine Polk Failla
United States District Court for the Southern District of New York
40 Foley Square, Room 618
New York, New York 10007

Re: *Dow Jones & Co. v. Perplexity AI, Inc.*, No. 24-cv-07984-KPF

Dear Judge Failla:

On behalf of Defendant Perplexity AI, Inc., I write to oppose Plaintiffs' April 20, 2026 letter seeking to compel an additional production of source code version history (*i.e.*, "commit history"). This requested production is substantively no different from the source code history Perplexity has already produced.

Neither Federal Rule of Civil Procedure 34 nor this Court's ESI or source code inspection orders (ECF Nos. 63, 93) entitle Plaintiffs to their preferred format of a commit history that was already provided to them "in a form or forms in which it is ordinarily maintained or in a reasonably usable form." Fed. R. Civ. P. 34(b)(2)(E)(ii). But that is what they are seeking. Perplexity has been responsive and cooperative with regard to Plaintiffs' source-code requests, and in fact has gone above and beyond what is required in that respect. This latest demand, however, has no merit, and should be rejected.

Plaintiffs do not dispute that Perplexity has provided the entire substantive code base for Perplexity Free, Perplexity Pro, Perplexity API, and Perplexity Bot, as agreed by the parties. In February 2026, Perplexity produced a complete Git clone of its [REDACTED] codebase with full version history; in March 2026, at Plaintiffs' request, Perplexity produced an additional [REDACTED] supplemental codebases comprising nearly [REDACTED] additional files across more than [REDACTED] subfolders; and on April 6, 2026, Perplexity produced an [REDACTED] JSONL file containing the complete commit history for all [REDACTED] source code repositories. Declaration of Jerry Ma ("Ma Decl.") ¶¶ 4-5, 8. The only dispute is whether Perplexity's production of the [REDACTED] JSONL file providing the commit history in a standard form is sufficient under Rule 34. It plainly is.

Commit history is "tantamount to source code itself," *IntelliCAD Tech. Consortium v. Suzhou Gstarsoft Co.*, 508 F. Supp. 3d 790, 794 (D. Or. 2020), in that commit logs "show details of changes made to source code." *Alexander Interactive, Inc. v. Adorama, Inc.*, 2013 WL 6283511, at *5 (S.D.N.Y. Dec. 4, 2013). To the extent Plaintiffs have a legitimate need to understand how Perplexity's source code evolved before and at the time suit was filed, the commit history is the record that answers exactly that question. Ma Decl. ¶¶ 9-16. The JSONL file Perplexity provided on April 6 reflects the complete commit history for all [REDACTED] repositories—showing, for each commit, when it was made, by whom, which files were changed, and the substance of those changes. *Id.* ¶¶ 13-14. Plaintiffs' own experts have already confirmed this data is accessible and usable: they successfully parsed and organized the JSONL file by repository themselves, without any assistance from Perplexity, and declined (through their silence)

Perplexity’s offer to enable additional tooling on the review computers that would have made their analysis even easier. Ma Decl. ¶¶16-18; email from J. Romeo to E. Pearson, Apr. 13, 2026. Plaintiffs cannot credibly maintain that data their own experts have already manipulated and organized is not in a “reasonably usable form.” *See* Fed. R. Civ. P. 34(b)(2)(E)(ii).

Despite these undisputed facts, Plaintiffs contend they are entitled to a second production *of the same commit history* because they would prefer it in a different format. Rule 34 expressly precludes Plaintiffs’ demand: “[a] party need not produce the same electronically stored information in more than one form.” Fed. R. Civ. P. 34(b)(2)(E)(iii). The JSONL file Perplexity produced is the native output format of GitHub’s Commits API—the industry-standard mechanism for exporting commit history at scale. Ma Decl. ¶9. It contains the complete commit history for all [REDACTED] repositories: [REDACTED]—precisely the information needed to trace how Perplexity’s systems functioned at any prior point in time. *Id.* ¶¶10, 13-14. The JSONL format is comparable in structure to other standard unified data formats—such as SQLite databases—that are conventionally produced as a single file and processed by recipients using standard tools. *Id.* ¶11. Plaintiffs are not entitled to compel Perplexity to produce the same information again in the form of Git repository clones simply because they would find that format more convenient.

Caselaw further confirms that, where, as here, relevant information can be derived from already-produced materials, the requesting party—not the responding party—bears the burden of doing the work to derive that information. *See, e.g., Nw. Biotherapeutics, Inc. v. Canaccord Genuity LLC*, No. 22 Civ. 10185 (GHW) (GS), 2026 WL 412081, at *3 (S.D.N.Y. Feb. 13, 2026) (“Courts have . . . declined to require a responding party to produce additional analyses that would otherwise be readily available to the requesting party based on factual material already produced.”). Courts have applied this principle directly to deny additional analysis beyond commit logs already made available for inspection. *See, e.g., dmarcian, Inc. v. DMARC Advisor BV*, No. 1:21-cv-00067-MR, 2024 WL 489312, at *3 (W.D.N.C. Feb. 7, 2024) (identifying differences between code versions via commit logs already produced for inspection satisfied discovery obligations).

Dr. Krein’s complaints do not compel deviation from these well-settled principles and, in any event, are overstated and/or mischaracterizations. For example, he mischaracterizes the JSONL file as [REDACTED] based on three categories of allegedly missing but in any event irrelevant, data: [REDACTED]

[REDACTED] ECF No. 118 ¶¶10, 13. As to [REDACTED] and their absence from the JSONL data does not impair analysis of how Perplexity’s text-processing, content-acquisition, or RAG systems functioned during the relevant period. Ma Decl. ¶20. As to the allegedly missing [REDACTED] Dr. Krein identifies, such [REDACTED] have no bearing on the analysis of Perplexity’s code history relevant to Plaintiffs’ claims; moreover, even a full Git repository clone would not have included the [REDACTED] data Dr. Krein says is missing. *Id.* ¶¶21-22. Finally, as to API truncation of “large changes,” [REDACTED] show any signs of truncation, and Perplexity’s review confirms those [REDACTED] one of which bear on Plaintiffs’ claims. *Id.* ¶23.

The JSONL commit history Perplexity has already provided fully serves Plaintiffs’ stated need to understand how Perplexity’s systems functioned before and at the time suit was filed: it records every change made to the relevant code, by whom, when, and to which files across all [REDACTED] repositories, providing precisely the historical record Plaintiffs claim they need without requiring the complete file-

by-file snapshot reconstruction Dr. Krein demands. As to Dr. Krein’s claim he has “never been asked to reconstruct whole source code versions based on a changelog,” that may be true on the specific facts of his prior engagements, but if commit logs were inadequate to understanding how source code evolved over time (which they are not), they would also be of no value as a discovery instrument. That is not the case, as courts have found. *See, e.g., dmarcian*, 2024 WL 489312, at *3.

Plaintiffs’ cited authority does not support their unorthodox request for duplicative discovery. In *Allscripts Healthcare, LLC v. Andor Health, LLC*, 2022 WL 1013624, at *3 (D. Del. Apr. 4, 2022), the plaintiffs’ expert testimony that the text file at issue there was “essentially useless” was not contradicted by any evidence. Here, in contrast, the declaration of Perplexity’s Deputy Chief Technology Officer rebuts Dr. Krein’s technical characterizations and explains in detail why the JSONL data is sufficient for the analytical purposes at issue. Ma Decl. ¶¶ 9-16. And unlike the text file provided in *Allscripts* that the plaintiffs claimed was not reasonably reviewable, the JSONL file Perplexity provided is not a raw text dump. It is a structured, machine-readable format natively produced by the GitHub API, as confirmed by the fact that Plaintiffs’ experts have already successfully parsed it. Ma Decl. ¶¶ 9-10, 17-18.

Even if the full Git clones Plaintiffs demand would provide some marginal benefit over the JSONL data Perplexity has already produced (which Perplexity disputes), the burden of producing [REDACTED] full repository clones would not be proportional to that benefit. *See* Fed. R. Civ. P. 26(b)(1). Producing Git bundle exports for [REDACTED] repositories is categorically more complex than the [REDACTED] Perplexity produced initially. Ma Decl. ¶¶ 30-31. Preparing each such export requires multiple steps (*e.g.*, mirror cloning, integrity verification, and bundle creation) requiring hands-on oversight by a senior engineer with the specific familiarity necessary to ensure completeness and accuracy. Ma Decl. ¶¶ 30, 32-34. Perplexity produced the [REDACTED] additional repositories Plaintiffs requested in March 2026 using GitHub’s API zipball format precisely because manually cloning repositories was not reasonably feasible at that scale. *Id.* ¶¶ 6, 31. Dr. Krein’s speculation that the job could be done in “[REDACTED] by a single professional engineer” does not account for these real-world constraints discussed above (which themselves do not account for these engineers’ competing responsibilities of building, testing, and maintaining Perplexity’s actual products). *Id.* ¶¶ 34-35. Realistically, the exports Plaintiffs demand would take approximately [REDACTED] in engineering time, without factoring in rate limit delays, to complete. *Id.* ¶ 36.

Finally, although it is unnecessary to the disposition of Plaintiffs’ letter request, which should be denied for the reasons stated above, Perplexity disputes Plaintiffs’ tacit premise that source code’s supposed relevance to Plaintiffs’ claims in this case justifies any amount of burden on Perplexity. Despite having (at minimum) weeks or months of access to both Log Database snapshots and Perplexity’s RAG database, not to mention over a year post-filing to continue investigation through their own querying of Perplexity’s answer engine, it appears that Plaintiffs have in fact found little or no evidence of any extensive (or other) reproduction of their works in these databases. In meet and confer discussions, Plaintiffs have repeatedly confirmed that despite having asserted in their complaint and amended complaint infringement of *all* of their copyrighted works, they intend to remove works for which they find no evidence of infringement in discovery. They are not finding such evidence. But rather than admitting the inevitable—that their theory of the case is in fact mistaken—Plaintiffs have doubled down, seeking more snapshots, more source code, and now commit history in different formats. Enough is enough: At some point, Perplexity’s discovery burden should be measured not against what Plaintiffs alleged in the first instance but against what Plaintiffs had some Rule 11 basis to allege, which at best appears to be a far smaller universe of even plausibly alleged infringements. For all of these reasons, the Court should deny Plaintiffs’ letter request.

Respectfully submitted,

/s/ Andrew H. Schapiro

Andrew H. Schapiro
Counsel for Defendant Perplexity AI, Inc.

cc: All counsel of record (via ECF)

The Court has reviewed Plaintiffs' motion to compel production of all missing Git Repositories (Dkt. #117, 119), as well as Defendant's above opposition (Dkt. #123, 127). The Court also thanks the parties for the helpful declarations of Dr. Krein (Dkt. #118, 120) and Mr. Ma (Dkt. #124, 126). After careful consideration of the relevant materials, Plaintiffs' motion to compel is DENIED.

In short, Plaintiffs have not persuaded the Court either (i) that they cannot get the information they need from the text file that Defendant has provided or (ii) that producing the Git Repositories is as easy as Dr. Krein says. Instead, the Court finds Mr. Ma's declaration more persuasive on both these points.

Specifically, Mr. Ma has explained that the JSONL file is a standard format for API data, and he has carefully laid out how Plaintiffs could analyze the file to find all the information they seek. (Dkt. #124 ¶¶ 9-16, 26-29). Additionally, he has effectively contested Dr. Krein's point about the burden of production by explaining how complying with Plaintiffs' request would require at least three work weeks rather than one to two days. (*Id.* ¶¶ 30-36). This case is thus different than *Allscripts Healthcare, LLC v. Andor Health, LLC*, No. 21 Civ. 704 (MAK), 2022 WL 1013624, at *3 (D. Del. Apr. 4, 2022), in which the Special Master granted a similar motion to compel because the "[d]efendants ... proffered no counter-declaration" to contest the plaintiff's arguments regarding the burdens that each side faced. Here, given the completeness and persuasiveness of Mr. Ma's declaration, the Court is unable to grant Plaintiffs' motion to compel on this record.

The Clerk of Court is directed to terminate the pending motion at docket entry 117.

SO ORDERED.

Dated: April 24, 2026
New York, New York



HON. KATHERINE POLK FAILLA
UNITED STATES DISTRICT JUDGE