

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

DOW JONES & COMPANY, INC.
and NYP HOLDINGS, INC.,

Plaintiffs,

v.

PERPLEXITY AI, INC.,

Defendant.

Civil Action No. 24-cv-7984

COMPLAINT

JURY TRIAL DEMANDED

Plaintiffs Dow Jones & Company, Inc. (“Dow Jones”) and NYP Holdings, Inc. (“NYP Holdings”) (collectively, “Plaintiffs”), by and through their attorneys, Torridon Law PLLC, for their Complaint, hereby allege against Defendant Perplexity AI, Inc. (“Perplexity” or “Defendant”), as follows:

NATURE OF THE ACTION

1. Perplexity is a generative artificial intelligence company that claims to provide its users accurate and up-to-date news and information in a platform that, in Perplexity’s own words, allows users to “Skip the Links” to original publishers’ websites.¹ Perplexity attempts to accomplish this by engaging in a massive amount of illegal copying of publishers’ copyrighted works and diverting customers and critical revenues away from those copyright holders. This suit is brought by news publishers who seek redress for Perplexity’s brazen scheme to compete for readers while simultaneously freeriding on the valuable content the publishers produce.

¹ On its website under the heading “Why Choose Perplexity?,” the very first reason given is “Skip the Links.” *What is Perplexity?*, Perplexity AI, <https://perma.cc/Q4VM-DYUJ>.

2. Dow Jones, NYP Holdings, and their corporate parent News Corporation (“News Corp”) are among the best-known news publishers in the world. Plaintiffs’ publications – including *The Wall Street Journal* and the *New York Post*, two of the most widely circulated newspapers in the United States – rely on the effort, talent, skills, and experience of accomplished journalists, editors, and other professional staff who meticulously investigate and skillfully write news stories, often under tight deadlines and in unpredictable circumstances. At times these journalists risk their lives and liberty to investigate, write, and break news stories that they believe will change the world for the better. These journalists, editors, and staff work hard to earn a living in the extremely competitive news industry. They do so by creating high-quality journalism – from the analytical, to the groundbreaking, to the provocative – to teach, to inform, and to entertain. The success of their work depends as much on uncovering the stories as it does on using their talent and skills to present the stories in a way that captures attention, spurs thinking, and provokes the imagination.

3. There is high demand for such content, particularly for accurate and up-to-date information presented in an easily digestible format. In the current digital age, Plaintiffs’ revenues for their original content come predominantly from selling subscriptions to their digital publications and from online advertising that is presented when a consumer visits the news publishers’ website directly or clicks on a search-engine or other link that brings the consumer to the website. Plaintiffs and other publishers also make significant revenue from licensing their content to others, including, but not limited to, artificial intelligence (“AI”) companies that compete with Perplexity.

4. Seeking to capitalize on this vast market for content, Defendant Perplexity has raised significant capital to build a so-called “answer engine.” Its AI “answer engine” copies on a massive scale, among other things, copyrighted news content, analysis, and opinion as inputs into

its internal database. It then uses that copyrighted content to generate responses to users' queries that are intended to and do act as a substitute for news and other information websites. Perplexity loudly touts that its answers to user queries are so reliable that its users can "Skip the Links" to the original publishers and instead rely wholly on Perplexity for their news and analysis needs. What Perplexity does not tout is that its core business model involves engaging in massive freeriding on Plaintiffs' protected content to compete against Plaintiffs for the engagement of the same news-consuming audience, and in turn to deprive Plaintiffs of critical revenue sources.

5. To build its substitute product, upon information and belief, Perplexity accesses and copies, without authorization or remuneration, vast numbers of webpages containing copyrighted material, including from Plaintiffs' webpages and third-party webpages containing Plaintiffs' licensed copyrighted works. In its quest to provide "answers," Perplexity accesses content from original sources, makes copies of that content that it includes in a massive "retrieval-augmented generation" ("RAG") database (called for purposes of this Complaint a "RAG index"),² provides that information from the index to a large language model ("LLM" or "model") (a form of artificial intelligence that functions as a complex predictive text machine), and repackages the original, indexed content in written responses ("outputs") to users. These outputs (or, in Perplexity's lingo, "answers") are machine-generated reproductions of human-created content arranged by LLMs and other tools that summarize and paraphrase original, human-generated content, even at times reproducing that content *verbatim*.

² For purposes of simplicity throughout this Complaint, Plaintiffs refer to Perplexity as having created a single RAG index. Upon information and belief, Perplexity may utilize other and/or additional methods of storing the content that Perplexity wrongfully takes from Plaintiffs for the purpose of providing a substitute product.

6. Perplexity’s conduct violates Plaintiffs’ exclusive rights under the Copyright Act in various ways. First, Perplexity’s actions at the *input* stage – copying without authorization massive amounts of Plaintiffs’ copyrighted works for inclusion into Perplexity’s RAG index (“inputs”) – constitute completed and massive copyright violations without even a colorable fair use.

7. The illegality of this massive copyright violation at the input stage does not depend on whether the particular outputs of Perplexity’s so-called “answer engine” are sufficiently similar in each instance to the copyrighted works of Plaintiffs as to constitute identical reproductions of those works. It is sufficient that Perplexity makes copies of Plaintiffs’ works on a grand scale to create reproductions and/or derivative content that is designed as a substitute for Plaintiffs’ works.

8. Nevertheless, the *outputs* of Perplexity’s products also unlawfully infringe on Plaintiffs’ copyrights in several ways. Perplexity’s “answers” to users’ queries often include full or partial verbatim reproductions of Plaintiffs’ news, analysis, and opinion articles. And worse, users can access verbatim reproductions of Plaintiffs’ content more frequently by purchasing a subscription to Perplexity’s premium service, “Perplexity Pro.” Other times, Perplexity turns Plaintiffs’ copyrighted articles into paraphrases or summaries of those copyrighted works that similarly serve as substitutes for accessing Plaintiffs’ copyrighted works on Plaintiffs’ own websites and/or licensed websites. The use of Plaintiffs’ copyrighted content to generate any such substitutes is not a fair use.

9. In addition to using Plaintiffs’ copyrighted work to develop a substitute product that reproduces or imitates Plaintiffs’ original content, Perplexity also harms Plaintiffs’ brands by falsely attributing to Plaintiffs certain content that Plaintiffs never wrote or published. Not infrequently, if Perplexity is asked about what Plaintiffs’ publications reported, Perplexity “answers” with false information. AI developers euphemistically call these factually incorrect

outputs “hallucinations.” Perplexity’s hallucinations can *falsely attribute* facts and analysis to content producers like Plaintiffs, sometimes citing an incorrect source, and other times simply inventing and attributing to Plaintiffs fabricated news stories.

10. Plaintiffs’ famous trademarks for *The Wall Street Journal* and the *New York Post*, whose enterprise and brand values are tied to their reputation for accuracy, are damaged and diluted by Perplexity’s use of Plaintiffs’ trademarks in connection with the hallucinations conjured by Perplexity’s “answer engine,” which deceives the public into believing that the hallucinations are authentic news stories published by Plaintiffs.

11. Perplexity’s business is fundamentally distinct from that of traditional search engines that also copy a vast amount of content into their indices but do so merely to provide links to the originating sites. In its traditional form, a search engine is a tool for discovery, pointing searchers to websites such as the pages of *The Wall Street Journal* or the *New York Post*, where the users can click to find the information and answers they seek. Those clicks in turn provide revenue for content producers. In part because traditional search engines that simply provide hyperlinks promote merely the *discovery* of copyrighted content, and not its *substitution* (and commercial monetization of that substitute), copyright holders historically have not sought to prevent or legally challenge the copying by traditional search engine companies. But unlike the business model of a traditional internet search engine, Perplexity’s business model does not drive business toward content creators. To the contrary, it *usurps* content creators’ monetization opportunities for itself.

12. Generative AI technology has enormous potential to benefit society, but neither society nor AI companies will benefit if original content creators are harmed. Undermining the financial incentives to create original content will result in less content being generated and/or less quality content being generated, which will also reduce the amount of content available to power

AI's growth. Fortunately, this is not a zero-sum game. Generative AI companies and content producers can mutually benefit from working together, and Plaintiffs welcome the opportunity to help develop and refine new generative AI tools.

13. Months ago, in July 2024, Plaintiffs sent a letter to Perplexity putting it on notice of the legal issues raised by Perplexity's unauthorized use of Plaintiffs' copyrighted works and offering to discuss a potential licensing deal. Perplexity did not bother to respond.

14. Other AI companies have engaged with Plaintiffs and other publishers, resulting in legitimate market-based licensing solutions. Indeed, as *Forbes* has recently observed, "AI content licensing initiatives abound."³ For example, News Corp recently partnered with OpenAI to license its content for certain uses in OpenAI's applications. OpenAI users will have the benefit of accessing Plaintiffs' content, whether quoted or summarized by OpenAI. This cooperative relationship will allow OpenAI and Plaintiffs to experiment with new product experiences and revenue models.

15. Generative AI technology can be developed in two ways. It can be developed legally by recognizing the legitimate rights of copyright holders and by including in the AI business model the legitimate costs and benefits of licensing the copyrighted material, or it can be developed illegally by stealing copyrighted material. Perplexity has elected the illegal approach, building a user base through theft of a massive volume of copyrighted material, perhaps because it believes that assembling prestige and market power will help bolster its bargaining position once it is called to account. Whatever its strategy, it is now time for that reckoning.

³ Bill Rosenblatt, *The Media Industry's Race To License Content For AI*, FORBES (July 18, 2024), <https://www.forbes.com/sites/billrosenblatt/2024/07/18/the-media-industrys-race-to-license-content-for-ai>.

THE PARTIES

16. Plaintiff Dow Jones & Company, Inc. (“Dow Jones”) is a Delaware corporation with its headquarters and principal place of business at 1211 Avenue of the Americas, New York, New York 10036.

17. Plaintiff NYP Holdings, Inc. (“NYP Holdings”) is a Delaware corporation with its headquarters and principal place of business at 1211 Avenue of the Americas, New York, New York 10036.

18. Defendant Perplexity AI, Inc. (“Perplexity”) is a Delaware corporation with its principal place of business at 575 Market Street, San Francisco, California 94105.

JURISDICTION AND VENUE

19. This civil action seeks damages, injunctive relief, and other equitable relief under the Copyright Act of 1976, 17 U.S.C. § 101 *et seq.*, and the Lanham Act, 15 U.S.C. § 1051 *et seq.*

20. This Court has jurisdiction over this civil action under 28 U.S.C. §§ 1331 and 1338.

21. Venue is proper in this Judicial District under 28 U.S.C. §§ 1391 and 1400(a).

22. This is the District in which a substantial portion of the events giving rise to the claims occurred and/or in which Plaintiffs’ injuries were suffered. Plaintiffs’ principal places of business and headquarters are in this District, a significant number of Plaintiffs’ customers are located in this State and District, and a significant share of Plaintiffs’ business from subscribers, website users, and licensing is transacted with individuals and entities residing in this State and District. As such, the injuries alleged herein from Perplexity’s infringement and other unlawful conduct foreseeably occurred in this State and District. In addition, Perplexity or its agents reside in this District and may be found in this State and District.

23. Defendant Perplexity is subject to the jurisdiction of this Court pursuant to N.Y. C.P.L.R. § 302(a)(1) and (3) as it has purposefully directed its activities at New York and has

purposefully availed itself of the benefits of doing business in New York, including by registering to transact and by transacting business in the State and District, and by supplying goods or services in the State and District, through both the promotion and distribution of its products and the sale of subscription services to customers in this State and District.

24. Perplexity has established a continuing presence in this State and District, including because Perplexity possesses or utilizes real property situated within the State and District and has employees in this State and District. Those employees include its co-founder and Chief Strategy Officer, as well as engineering, infrastructure, and content staff, who are actively engaged in developing, implementing, and promoting the technology that Perplexity uses to gather and misappropriate Plaintiffs' copyrighted content.

25. Perplexity aggressively promotes and advertises its products and services in this State and District, including through its dynamic, heavily interactive website (<https://www.perplexity.ai> and associated subpages) and mobile applications. Perplexity's website targets customers in this State and District with promotional material tailored to a New York audience, including a web page inviting visitors to "Discover New York with Perplexity."⁴ Perplexity's marketing activities include promoting on its Instagram account a massive billboard in Times Square from September 2024 which read "Congratulations Perplexity on 250 million questions answered last month."⁵

⁴ *Discover New York with Perplexity*, Perplexity AI (last visited Oct. 17, 2024), <https://www.perplexity.ai/encyclopedia/discovernewyork>.

⁵ *@perplexity.ai, Instagram* (Sept. 4, 2024), https://www.instagram.com/perplexity.ai/p/C_g2TonSHC5.

26. Perplexity processes subscription purchases from customers in this State and District, transmits Plaintiffs' copyrighted content to users in this State and District, and has a significant number of customers in this State and District.

27. As a direct and proximate result of Perplexity's unauthorized use and/or dissemination of Plaintiffs' copyrighted works and trademarks in New York and elsewhere, Plaintiffs have lost and will continue to lose revenue and profits from the market for content licensing, subscribers, visitors, and users.

FACTUAL ALLEGATIONS

I. Plaintiffs' Robust Businesses and Copyrighted Works

28. Dow Jones began in 1882 as a niche news agency in a Wall Street basement, founded by reporters Charles Dow, Edward Jones, and Charles Bergstresser. Publishing the first edition of *The Wall Street Journal* in July 1889, Dow Jones has now expanded into a worldwide news powerhouse. It creates and distributes some of the most widely recognized and reputable publications in the news industry, including, in addition to *The Wall Street Journal*, Dow Jones Newswires, MarketWatch, *Financial News*, and *Barron's*.

29. Dow Jones is a trusted source of accurate, original news stories, data and analytics, and financial and business insight for millions of customers across the country and around the world.

30. A recipient of 39 Pulitzer Prizes, the award-winning newsroom at *The Wall Street Journal* sent one of the first news reports of a plane crashing into the World Trade Center on September 11, 2001, broke the story of fraud and corruption at Theranos which led to criminal prosecutions and convictions, uncovered thousands of federal officials having investments in companies that stood to benefit from their agencies' work, and first revealed a social media

company's failure to take remedial measures despite internal knowledge of its platforms' ill effects on minors.

31. Dow Jones owns copyrighted content in *The Wall Street Journal*, and registers copyrights with the United States Copyright Office for *The Wall Street Journal*. Upon information and belief, Dow Jones's registered copyrighted works that Perplexity has been able to access electronically have been copied and/or reproduced by Perplexity into its RAG index, including the content covered by the registrations listed at Appendix 1.

32. Dow Jones is also the owner of a number of federally registered trademarks for the mark *The Wall Street Journal* and the mark *WSJ* including Registration Nos. 408,379, 1,960,159, 4,297,319, 4,297,321, 4,298,327, 3,981,663, and 4,813,188, and others, which were registered on August 8, 1944, March 5, 1996, March 5, 2013, March 5, 2013, March 5, 2013, June 21, 2011, and September 15, 2015, respectively, and by virtue of their continued use are incontestable.

33. NYP Holdings is the parent company of the *New York Post*, a trusted major news source for millions of readers and ranking last year as the third largest newspaper by print circulation in the United States. The paper was first published in 1801 by Alexander Hamilton. It is America's oldest continuously published newspaper and has one of the most visited news websites in the country.

34. NYP Holdings owns copyrighted content in the *New York Post*, and registers copyrights with the United States Copyright Office for the *New York Post*. Upon information and belief, NYP Holdings' registered copyrighted works that Perplexity has been able to access electronically have been copied and/or reproduced by Perplexity into its RAG database, including the content covered by the registrations listed at Appendix 2.

35. NYP Holdings is also the owner of a number of federally registered trademarks for the mark *New York Post* including Registration Nos. 1,526,818, 2,213,076, and 3,639,502, and others, which were registered on February 28, 1989, December 22, 1998, and June 16, 2009, respectively and by virtue of their continued use are incontestable.

36. Providing comprehensive, compelling news coverage, analysis, and opinion requires the immense innovation and creativity of many contributors. Journalists at Dow Jones and NYP Holdings decide not only what stories to tell but how to tell them, making original choices about what material should be included or emphasized, how a story should be organized and framed, and in what style or voice the story should be presented. Journalistic writing often requires significant ingenuity as journalists must present compelling narratives to audiences within the confines of fact and truth. The success and survival of news publishers depend just as heavily on which stories they choose to uncover and tell as on *how* they tell those stories.

37. News publishing, particularly breaking daily news, also has a significant financial cost and risk. News publishers like Dow Jones and NYP Holdings employ journalists, subject-matter experts, photographers, videographers, editors, and administrative staff. They maintain extensive equipment and offices, including global networks of correspondents and bureaus, to cover events around the world. And stories require a multilayered editing process to ensure standards of quality and reliability, which further adds to the cost of creation.

38. The cost of journalism is also much more than financial. Journalists may risk their lives and liberty in furtherance of their commitment to an informed public. They report from war zones, investigate terror networks, and report on major natural disasters. They grapple with arrest, incarceration, and wrongful prosecution, and sometimes even tragically lose their lives.

39. News publishers can support their journalists' work and invest skill, creativity, time, and money into the important endeavor of gathering and presenting the news only if there is a market that fairly compensates them for these efforts.

40. Before widespread internet access, news publishers funded their operations entirely from subscriptions, newsstand sales, and advertisements in their printed newspapers. But the economics of newsgathering have significantly changed in the past quarter-century, with many news companies downsizing or ceasing production of print editions altogether. Most now rely overwhelmingly on digital content, including paid access to copyrighted content and content supported by advertising revenue. Those revenues depend on consumers subscribing to the publications and, in the case of advertising-supported content, clicking on the links to the publisher's own webpages and to licensed webpages. Any company that copies news publishers' original works to enable and encourage consumers to "Skip the Links," as Perplexity urges, purposefully diverts readers' attention to itself and away from the original news publishers and siphons off for itself the revenue opportunities that support news and journalists.

41. Dow Jones, NYP Holdings, and other publishers have entered into agreements expected to yield significant revenue through the licensing of their copyrighted articles to AI companies that want to utilize popular, high-quality, human-created journalism for use by the companies' AI applications. Revenue received from legitimately-run AI companies supports the costs of news gathering. This revenue also establishes that there is a market for the licensing of human-generated content for lawful use in AI technologies.

42. Plaintiffs' content is highly valued in this market.

II. Perplexity's Business, Generative AI Technology, and "RAG" Index

43. Perplexity is a generative AI company.

44. Generative AI systems can generate content that mimics natural language in response to a prompt using technology called large language models.

45. LLMs function as complex predictive text machines. They assign different words a different number, or "vector," that represents their relationship to other words. The model can produce text that appears like natural language and/or language resembling the style of a certain author or publication.

46. LLMs are "trained" on large amounts of content that enable them successfully to assemble sentences and paragraphs for a reader to comprehend. News articles, analysis, and editorials serve as very useful content in this training due to, among other things, their clarity and structure, the editing and quality control they receive, their wide range of topics, their range of perspectives, the recency of their information, their writing style, and their tone.

47. After an LLM is initially trained on vast amounts of content containing many natural language expressions, the model can be "fine-tuned." Fine-tuning is a process by which an LLM is trained further on a smaller, more specific set of data that allows the model to become more specialized in conducting a specific task, responding to prompts specific to a subject area, or recognizing nuances in particular questions.

48. Fine-tuning might also ensure that the LLM responds to certain prompts by mimicking a certain linguistic style. For example, outputting a cooking recipe requires a distinct output style from recounting the statistics of the Allies' landing at Normandy's beaches on D-Day, or from writing a poem about the summer wind. A medical treatise uses a distinct linguistic style from a sports recap.

49. To improve a model’s chances of presenting factually correct outputs, generative AI companies wishing to offer products that produce current and reliable information often incorporate an additional step in their process after training and fine-tuning. This additional and important step is known as “retrieval-augmented generation,” shortened as “RAG,” or known colloquially as “grounding.”

50. RAG enables AI applications to focus their models on a very specific, limited set of data or piece of content in response to a particular user’s query. RAG indices also give AI applications the ability to collect and store new content created by authors after the LLMs they utilize have been trained. To employ a RAG system, AI applications typically utilize indexed databases that house all the content from which the AI application will retrieve specific information to generate outputs for its users. The larger the index, the more “answers” the AI application can provide.

51. In Perplexity’s words, it “scours the internet, gathering information from authoritative sources like articles, websites, and journals.”⁶ It then, “compiles the most relevant insights into a coherent, easy-to-understand answer” automatically generated from those original sources.⁷

52. The assembling of authoritative sources for a RAG index is a distinct process from assembling content for training an LLM and from assembling content for fine-tuning an LLM. RAG indices are specifically comprised of content that AI companies want to use as source material from which to generate the “answers” to user prompts and questions. They consist *only* of those high-quality sources that AI companies deem trustworthy – the most trustworthy sources

⁶ How does Perplexity work?, Perplexity AI, <https://perma.cc/4F9F-R7PS>.

⁷ *Id.*

being those that expend the most human capital on drafting, fact-checking, editing, and reviewing the content.

53. RAG indices are organized in a manner to allow rapid-retrieval, and they are updated continuously by web crawlers that scrape and add the most current content to the RAG index. Rapid retrieval is essential to generative AI technology's ability to avoid latency in producing outputs. In particular, a robust and up-to-date RAG index that includes high-quality, human-generated content, such as news, analysis and opinion stories from Dow Jones and NYP Holdings, is essential to Perplexity's AI products. As Perplexity cofounder and current CEO Aravind Srinivas has acknowledged, content from major news publications is "generally more reliable than Substack posts, which may be more opinionated."⁸

54. Upon information and belief, since it was launched in 2022, Perplexity has copied, or caused to be copied, all of the copyrighted content of Dow Jones and NYP Holdings into its RAG index that it has been able to access, including content covered by the registrations listed at Appendix 1 and Appendix 2.

55. There are established, growing markets both for content that can be copied into a RAG index as well as for RAG indices themselves, which can be bought, sold, and licensed.

56. Once original content from the RAG index is selected from which to derive an output, the content is copied again into the "context window" of an LLM, along with the user's query. The LLM then uses advanced predictive text technology to reproduce or generate a derivative of the original content in an "output" of sentences and phrases that respond to the user's query.

⁸ See, e.g., Joanne Chen, *How Perplexity.ai Is Pioneering The Future Of Search*, FORBES (Sept. 6, 2023), <https://www.forbes.com/sites/joannechen/2023/09/06/how-perplexityai-is-pioneering-the-future-of-search> (referencing *The New York Times*).

57. For example, if asked to provide three take-aways from a magazine or news article, the AI application will first identify the article it wants to utilize for its response. Then it will feed the article, along with the prompt, to the LLM. The LLM will then conduct a multi-step process by which it identifies groupings of words likely to reflect a few different points in the article and provides that content to the user in a readable format, whether in verbatim text, a paraphrased summary, or in some combination of the two.

58. Perplexity has explained that a distinct feature of its AI application is that its RAG technology tells the LLM exactly which original content to turn into its “answer.” Perplexity believes it can produce more accurate answers because it relies so heavily – and sometimes exclusively – on sources it deems particularly trustworthy. According to Srinivas, Perplexity’s principle is: “given a query, always retrieve relevant documents and pick relevant paragraphs from each document and use those documents and paragraphs to write your answer for that query. The principle in Perplexity is you’re not supposed to say anything that you don’t retrieve.”⁹

59. For several reasons, including because Perplexity generates outputs for users on an individualized basis, it is effectively impossible for original content creators to uncover all of the ways in which their copyrighted works are being repackaged and sold to Perplexity’s users, short of legal proceedings compelling disclosures by Perplexity.

60. Upon information and belief, Perplexity has opted to ignore copyright law in an attempt to quickly get its products to market and to acquire a sizable customer base ahead of other

⁹ Lex Fridman Podcast, *Aravind Srinivas: Perplexity CEO on Future of AI, Search & the Internet*, YOUTUBE, at 01:56:50 (June 19, 2024), <https://www.youtube.com/watch?v=e-gwvmhyU7A>.

AI companies. In the words of its CEO Aravind Srinivas, “Once you have the user love [your technology] . . . it’s very hard to lose from there.”¹⁰

III. Perplexity’s Response to Public Criticism That It Ignores Copyright Law

61. Perplexity has received significant public criticism for ignoring copyright law and otherwise acting unethically in its race to get users.¹¹ At various times, Perplexity has responded publicly with a series of nonresponsive, inaccurate, and/or contradictory excuses.

62. Initially, Perplexity attempted to justify its massive illegal copying of copyrighted material by touting repeatedly its “Cited Sources” feature, as though a citation to an original source with a link that Perplexity actively encourages users to “skip” would make up for stealing copyrighted material.

63. When asked at a technology conference how he will “deal with copyright issues . . . [because he] definitely will take data from the internet from some random website and the [] traffic will be reduced from their website,” Srinivas responded: “The way we are thinking about it is we are at least attributing every part of the answer. Where are we getting it from in terms of inline citations as well as the source panel at the top.”¹²

¹⁰ Outset Capital, *Perplexity CEO Aravind Srinivas Thursday Nights in AI*, YOUTUBE, at 12:55 (July 18, 2023), <https://www.youtube.com/watch?v=jksGQhMtXjo>.

¹¹ See, e.g., Elizabeth Lopatto, *Perplexity’s Grand Theft AI*, THE VERGE (June 27, 2024), <https://www.theverge.com/2024/6/27/24187405/perplexity-ai-twitter-lie-plagiarism> (describing Perplexity as a “rent-seeking middleman on high-quality sources” that “starve[s] the primary source of ad revenue”); Dhruv Mehrotra & Tim Marchman, *Perplexity Is a Bullsh*t Machine*, WIRED (June 19, 2024), <https://www.wired.com/story/perplexity-is-a-bullshit-machine> (discussing Perplexity’s reliance on recent news articles for its content as well as its tendency to falsely attribute information) (asterisk added); Casey Newton, *How to Stop Perplexity and save the web from bad AI*, PLATFORMER (June 20, 2024), <https://www.platformer.news/how-to-stop-perplexity-oreilly-ai-publishing> (examining Perplexity’s “ethical rather than technical” innovation in using protected content without permission).

¹² Indian Institute of Technology Madras, Office of Alumni and Corporate Relations, *Talk by Shri. Aravind Srinivas, Founding Story and Journey of Perplexity*, YOUTUBE, at 49:05 (Jan. 18, 2024), <https://www.youtube.com/watch?v=ygRVDIwheB4>.

64. This attempt to excuse copyright infringement by pointing to the inclusion of “cited sources” is a repeated theme in Perplexity’s public statements. Srinivas was asked a similar question by interviewer Andrew Ross Sorkin. Specifically, Mr. Sorkin asked: “There is a real question mark about what happens to that economy, not the search advertising part of the economy, but the actual economy of those who are creating the underlying content that effectively is being scraped. If it all goes back into a system like yours, and you produce the perfected result, and nobody ultimately is going back to the original source or website, what that does to the system. What do you think about that?”¹³

65. In response, Srinivas again obfuscated by referencing Perplexity’s cited sources feature, stating in part: “So Perplexity has been revolutionary in the fact that ever since the first day we launched, we always cite our sources.”¹⁴

66. In response to a *Forbes* exposé about Perplexity’s unauthorized use of its work, Srinivas disregarded *Forbes*’ primary concern, saying in part: “We agree with the feedback you’ve shared that it should be a lot easier *to find the contributing sources* and highlight them more prominently.”¹⁵

67. Srinivas has also stated: “Citations are a very big thing in academia . . . Everyone feels proud if their paper gets cited by other people, and especially if it gets cited by people they

¹³ News Release, *CNBC Exclusive: CNBC Transcript: Perplexity Founder & CEO Aravind Srinivas Speaks with CNBC’s Andrew Ross Sorkin on “Squawk Box” Today*, CNBC (Apr. 23, 2024), <https://www.cnbc.com/2024/04/23/cnbc-exclusive-cnbc-transcript-perplexity-founder-ceo-aravind-srinivas-speaks-with-cnbc-andrew-ross-sorkin-on-squawk-box-today.html>.

¹⁴ *Id.*

¹⁵ @AravSrinivas, X (June 7, 2024 3:21 PM) <https://x.com/AravSrinivas/status/1799159732126794017> (emphasis added); see also Randall Lane, *Why Perplexity’s Cynical Theft Represents Everything That Could Go Wrong With AI*, FORBES (June 11, 2024), <https://www.forbes.com/sites/randalllane/2024/06/11/why-perplexitys-cynical-theft-represents-everything-that-could-go-wrong-with-ai>.

respect. They feel very good about it. And in our user interface, even though we give the answer, we do show the user exactly where the answer is coming from.”¹⁶

68. As Srinivas surely knows or should know, academic standards for avoiding plagiarism are wholly independent from copyright law.¹⁷ Dow Jones and NYP Holdings editors and journalists are not graduate students working out of a library or lab, eager to have someone acknowledge and utilize their research. They are professional journalists and publishers – working under high-pressure deadlines, sometimes in dangerous places – whose livelihoods depend on the enforcement and monetization of their intellectual property rights.

69. Upon information and belief, Perplexity’s citations make users *less* inclined to visit the original content source, because, as Perplexity has boasted, citations make content appear more “Reliable,”¹⁸ allowing Perplexity’s readers to feel more confident that they can “Skip the Links” when they believe they are reading content from credible sources.

70. This observation is corroborated by Plaintiffs’ experience of detecting virtually no click-through traffic on their websites from “cited sources” links on Perplexity, despite Perplexity receiving approximately 250 million queries per month.¹⁹

¹⁶ Indian Institute of Technology Madras, Office of Alumni and Corporate Relations, *Talk by Shri. Aravind Srinivas, Founding Story and Journey of Perplexity*, YOUTUBE, at 17:57 (Jan. 18, 2024), <https://www.youtube.com/watch?v=ygRVDIwheB4>.

¹⁷ See, e.g., *Avoiding Plagiarism Guide*, APA Style 7th Edition (last visited Aug. 30, 2024), <https://apastyle.apa.org/instructional-aids/avoiding-plagiarism.pdf>.

¹⁸ See *What is Perplexity?*, *supra* note 1 (promoting Perplexity’s “Reliable sources” with an explanation that “[e]very answer is backed by citations from trusted news outlets, academic papers, and established blogs”).

¹⁹ Madhumita Murgia & Cristina Criddle, *Perplexity’s popularity surges as AI search start-up takes on Google*, THE FINANCIAL TIMES (Aug. 9, 2024), <https://www.ft.com/content/87af3340-2611-4650-9ae3-036927e9f65c>.

71. Perplexity’s next attempt to deflect criticism and legal action against their massive illegal copying and use of copyrighted works was their announcement of a so-called “Publishers’ Program.”

72. Perplexity cofounder and current Chief Technology Officer Denis Yarats has essentially acknowledged the need to enter into agreements compensating the owners of the copyrighted works that Perplexity appropriates for its own commercial benefit. “I think there has to be probably something where you have to work together with the content creators,” he said.²⁰ “It’s going to be interesting to figure out what is the beneficial way for both of the parties because, yeah, I think the content generation is very important. And I think, you know, if people stop producing it, it’s then, yeah, probably not good for anybody.”²¹

73. According to Perplexity, the Publishers’ Program was developed “[t]o further support the vital work of media organizations and online creators,” because the company “need[s] to ensure publishers can thrive as Perplexity grows.”²² The program purports to share an unspecified portion of ad revenue from advertisements that Perplexity plans to host in “coming months.”²³

74. While Perplexity is correct in conceding that generative AI is unsustainable without a thriving market for publisher content, the Publishers’ Program is no solution or defense to Perplexity’s infringement. Rather, this program is Perplexity’s naked attempt to dictate unilaterally the terms of a license to owners of original content from whom Perplexity has already taken

²⁰ Gradient Dissent, *Transforming Search with Perplexity AI’s CTO Denis Yarats*, YOUTUBE, at 31:27 (June 20, 2024), <https://www.youtube.com/watch?v=gVP-DxqatLQ>.

²¹ *Id.*

²² *Introducing the Perplexity Publishers’ Program*, Perplexity AI (July 30, 2024), <https://www.perplexity.ai/hub/blog/introducing-the-perplexity-publishers-program>.

²³ *Id.*

copyrighted material. In the negotiation of a valid license to copyrighted material, an infringer does not unilaterally dictate the terms of the license to its victims.

75. At other times, Perplexity has publicly alluded to the possibility that its copying of publishers' copyrighted works could possibly be justified as a fair use. Responding to a question about Perplexity's scraping and reprinting of content, Perplexity's Chief Business Officer Dmitry Shevelenko is quoted as saying: "There's intricacies of fair use and copyright law where we feel we're kind of, you know, clearly within those bounds."²⁴

76. This half-hearted suggestion of a fair use defense is also incorrect. Perplexity's use is plainly and solely for a commercial purpose. Moreover, upon information and belief, it copies into its index every single word of Plaintiffs' copyrighted works that it can get its hands on. Additionally, the use to which it puts these copies is to create a commercial substitute for Plaintiffs' protected works – in Perplexity's own words, to allow and encourage users to "Skip the Links" to Plaintiffs' original works. Such substitution causes substantial harm to Plaintiffs' traditional advertising and subscription revenues. Perplexity's conduct also harms Plaintiffs' additional, established revenue stream from licensing to more scrupulous AI companies. Nor is Perplexity's use of Plaintiffs' copyrighted works transformative in the way that copyright law sometimes protects. A transformative work adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message. Here, all that Perplexity's so-called "answer engine" does is massively copy the copyrighted expression, meaning, and message of others and repackage it to the consumer. Perplexity's machine adds no new expression, meaning, or message of its own.

²⁴ Kylie Robison, *Perplexity is cutting checks to publishers following plagiarism accusations*, THE VERGE (July 30, 2024), <https://www.theverge.com/2024/7/30/24208979/perplexity-publishers-program-ad-revenue-sharing-ai-time-fortune-der-spiegel>.

IV. Perplexity's Massive Illegal Copying of Plaintiffs' Original Works as Inputs for Its RAG Index

77. The most basic copyright violation in this case is Perplexity's unlicensed and illegal copying of a massive amount of Plaintiffs' copyrighted works as inputs into its RAG index.

78. Perplexity engages in this wrongful conduct in order to deliver on its promise that its users can "Skip the Links" to the websites of original content creators.

79. Perplexity needs this original content to achieve what it purports to deliver to its users: "the equal in human labor of a journalist or an academic scholar, or a person writing Wikipedia pages, being done for you in a matter of seconds."²⁵

80. Perplexity has marketed itself as, among other things, a substitute purveyor of news, even though Perplexity does not employ any journalists or editors to produce original news stories. According to Perplexity's website, it seeks to attract customers interested in a "daily [news] briefing."²⁶

81. Upon information and belief, Perplexity has copied hundreds of thousands of Plaintiffs' copyrighted articles for its RAG database without authorization, the true extent of which is a fact uniquely in the possession of Perplexity.

82. Perplexity's need for a massive, unauthorized trove of Plaintiffs' content can be explained by its CEO's admission that "Perplexity would be useless" if there were no new web content for it to utilize.²⁷ As a purveyor of news and other news-related content, but without its

²⁵ Building the Metaverse, *AI and the Search for Truth and Answers*, YOUTUBE, at 02:48 (July 20, 2023), <http://www.youtube.com/watch?v=TTjEr7TFcmQ>.

²⁶ *What is Pro Search?*, Perplexity AI (last visited Aug. 30, 2024), <http://perplexity.ai/hub/faq/what-is-pro-search>.

²⁷ Mark Sullivan, *Perplexity CEO Aravind Srinivas responds to plagiarism and infringement accusations*, FAST COMPANY (June 21, 2024), www.fastcompany.com/91144894/perplexity-ai-ceo-aravind-srinivas-on-plagiarism-accusations.

own newsroom, Perplexity must copy content from the websites of “[r]eliable sources” and “trusted news outlets”²⁸ into its RAG index. As Shevelenko said, presumably referring to companies like Plaintiffs, “obviously larger publishers tend to produce the types of content that is more authoritative and has more verified facts.”²⁹

83. Upon information and belief, Perplexity copies copyrighted content in numerous ways, each of which is a copyright violation. Sometimes Perplexity copies Plaintiffs’ copyrighted works from Plaintiffs’ own websites by crawling and scraping them with its own crawlers. In doing so, sometimes Perplexity and/or its agents have purposely ignored and/or evaded technological features, such as the long accepted worldwide robots.txt protocol, a tool designed to instruct web crawlers *not* to copy digital content.³⁰ Other times, Perplexity crawls and scrapes Plaintiffs’ copyrighted content from third party websites of Plaintiffs’ licensees. Perplexity may also crawl and scrape content using a third-party crawler under its direction and control.

84. In all events, Perplexity’s massive copying and use of Plaintiffs’ copyrighted works into its RAG index is a completed violation of copyright law.

V. Perplexity’s Infringing Outputs

85. In addition to Perplexity’s massive and illegal copying of Plaintiffs’ copyrighted works as inputs into its RAG index, many of the *outputs* – answers – of Perplexity’s so-called “answer engine” constitute distinct copyright violations.

²⁸ *What is Perplexity?*, *supra* note 1.

²⁹ Charlotte Tobitt, *Perplexity to share ad revenue with signed-up publishers after flurry of criticism*, PRESS GAZETTE (July 30, 2024), <https://pressgazette.co.uk/news/perplexity-publishers-revenue-sharing>.

³⁰ *See generally* Lopatto, *supra* note 11 (discussing Perplexity’s use of scrapers that ignore the robots.txt protocol).

86. Sometimes, Perplexity’s answers contain full or partial verbatim reproductions of Plaintiffs’ copyrighted articles. Moreover, upon information and belief, these reproductions are more likely to be obtained by the user if the user subscribes to Perplexity’s “Perplexity Pro” for \$20 per month.

87. Other times, Perplexity’s outputs will be reworded into text resembling a paraphrase or summary that is sourced from Plaintiffs’ copyrighted content.

88. Upon information and belief, in each of these cases, these output answers are designed to, and do, act as substitutes for users clicking on links and otherwise going to Plaintiffs’ own or licensed third-party websites to read their copyrighted work, thus diverting revenue away from Plaintiffs to Perplexity.

89. For example, in the output shown below, Perplexity generates a reproduction of the full text of one of Plaintiffs’ copyrighted articles. When a Perplexity Pro user asked it to identify what the *New York Post*’s lead sports columnist recently said about the New York Mets, Perplexity described several pieces of content. When the user then asked Perplexity to generate the full text of one of the articles identified, Perplexity obliged, providing a verbatim reproduction of the article in full, which is registered with the U.S. Copyright Office as part of Registration No. TX 9-421-491. Perplexity’s output is shown in the image below, and a copy of the original *New York Post* article (which appeared on page 74 of the June 23, 2024 print edition) is attached in Appendix 3.

[output image on next page]

Perplexity provides a Perplexity Pro user the full text of a *New York Post* article

Can you provide the fultext of that article

Pro Search

- Search for the full text of Mike Vaccaro's article about his first baseball game at Shea Stadium ▼
- Fetch the full content of the article from the URL found in the search results ▼

Sources

FIRST LOVE

pressreader · 1

Trip to Shea Stadium 50 years ago started a lifelong love of baseball

nypost · 2

Mike Vaccaro: "Isiah Thomas is on the Back Page in a Full Clown..."

evergreenpodcas... · 3

View 5 more

Answer

The full text of Mike Vaccaro's article about his first baseball game at Shea Stadium 50 years ago is provided in the search results, specifically in sources ¹ and ². Here is the full text from source ¹:

FIRST LOVE

Trip to Shea 50 years ago started lifelong relationship with baseball

A FEW people reached out the past couple days, surprised that I didn't write a Father's Day column last week. Fact is, I wanted to save it a week for reasons you are about to see. And that decision, as you will also see, probably proved fortuitous.

I wanted to wait because next Saturday will mark the 50th anniversary of the first baseball game I ever attended in person. I was 7 years old. My father took me, of course. We did what we would do for almost every Mets game we ever attended: We parked the car at my grandmother's house on 51st Avenue in Corona, she made us a few sandwiches to go, and we hoofed it the couple of blocks to Shea Stadium.

It was June 29, 1974. I cannot tell you where my wallet is located right now. I'm iffy on the status of my iPhone. My car keys? Fifty-fifty. But I can recount — as if I'm looking at the home movie — every minute of that day, every second, warehoused in my memory.

From the moment Shea came into view, I'm not sure I blinked. The enormity of it, the odd orange-and-blue tiles randomly affixed to the exterior. And the blinding green of the grass inside. We took our seats. It was Old-Timers' Day at Shea, so we had gotten there early, so my dad started pointing out everything he thought I should know. "That's the dugout." "That's the bullpen." "That's the batting cage." "That's where they keep the tarp rolled up." "That's the press box." Here's one of the reasons it was important for me to write about that this week. My reply to that one was: "What's a press box?" He explained that's where the sportswriters went to write their stories about the game we were about to watch. And I swear this is true: My father commuted to the city from Long Island every day. He'd buy a home and hand me the tickets, and I would absorb every word.

+ Ask follow-up
Pro
↑

From the moment Shea came into view, I'm not sure I blinked. The enormity of it, the odd orange-and-blue tiles randomly affixed to the exterior. And the blinding green of the grass inside. We took our seats. It was Old-Timers' Day at Shea, so we had gotten there early, so my dad started pointing out everything he thought I should know. "That's the dugout." "That's the bullpen." "That's the batting cage." "That's where they keep the tarp rolled up." "That's the press box." Here's one of the reasons it was important for me to write about that this week. My reply to that one was: "What's a press box?" He explained that's where the sportswriters went to write their stories about the game we were about to watch. And I swear this is true: My father commuted to the city from Long Island every day. He'd buy the Daily News going in every morning, and The Post (an afternoon paper then) coming home every night. And starting the next Monday, and for the next 11 years, he would come home and hand me The Post, and I would absorb every word.

By that first Thursday — again, swear this is true — I announced at dinner: "I want to be a sports columnist for the New York Post when I grow up."

So when I occasionally mention that I have my dream job ... I am deadly serious. From when I was 7.

The other indelible thing that happened that day: At the very end of the introduction of the Old Timers, in from center field, together, walked Joe DiMaggio, Mickey Mantle, Duke Snider and Willie Mays. It was an electric moment. I'd never heard sound of any kind as intense as the roar that filled old Shea. But then I turned to my father and saw something I'd never seen before. He was crying. "I'm not sad," he said. "Just the opposite. These were the guys who made up most of my childhood."

Then he explained: Joe D. was his all-time favorite, and always would be, and he explained why. He lauded Mickey. He praised the Duke. "But Willie," he said, "was the best that ever played the game, in my mind. Best I ever saw in person anyway."

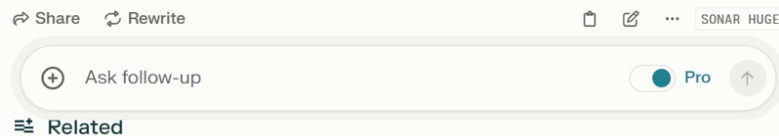
The game? Well, Cleon Jones and Wayne Garrett hit home runs, and the Mets won 4-0. Jon Matlack threw a one-hitter — the only hit a single by, of all people, the opposing pitcher, John Curtis.

"You have no idea what you almost just saw," my father said, shaking his head, as we walked back to my grandmother's house.

"I'm sure I'll see one eventually," I said.

(SPOILER ALERT: I have now attended about 5,000 baseball games in the past 50 years, give or take. I scrambled to the press box of the last inning of Johan Santana's no-no. I still haven't seen a full one. And it's getting late early out here.)

Fifty years. And it's funny. I wanted to include a picture of the ticket stub, which I still have, but it is permanently affixed in a scrapbook and there was no way to remove it without tearing it. The one pictured here looks just like it, I assure you. It'll have to do. You need to respect the most sacred artifacts, after all, at all costs.



90. This is merely one example of Perplexity or Perplexity Pro generating verbatim text from one of Plaintiffs' copyrighted articles, and upon information and belief, there are many others. Plaintiffs have observed that by using Perplexity Pro a user can generate between three full paragraphs to an entire article of text verbatim from the *New York Post*.

91. The next example shows Perplexity's free application providing an in-depth summary of a paywalled article of *The Wall Street Journal*, which is registered with the U.S. Copyright Office as part of Registration No. TX 9-429-903. Perplexity's output is shown in the image below, and a copy of *The Wall Street Journal* original article (which appeared on pages A1 and A8 of the July 31, 2024 print edition) is attached in Appendix 4.

Perplexity's detailed summary of a paywalled article from *The Wall Street Journal*

Summarize the WSJ article by those reporters that was published today

Updated just now

Sources

- Alistair MacDonald (@macdonaldajm) / X - 1
- Playbook PM: The pro-Harris ad blitz begins - POLITICO - 2
- Lara Seligman on X: "SCOOP: The U.S. has agreed to arm dozens ... - 3
- View 2 more

Answer

Based on the information provided in the search results, I can summarize the key points from the Wall Street Journal article by Lara Seligman and Alistair MacDonald published on July 30, 2024:

1. The United States has agreed to arm dozens of F-16 fighter jets being sent to Ukraine with American-made missiles and other advanced weapons ² ³ ⁵.
2. This decision addresses a previously unclear aspect of the F-16 transfer, as it wasn't certain where the missiles and munitions for these jets would come from ² ⁵.
3. The Pentagon made this decision despite having "limited inventory and production capability" for these advanced weapons systems ² ⁵.
4. The F-16 jets themselves are being provided by several European nations, while the U.S. will supply the advanced armaments to equip them ⁵.
5. This move represents a significant enhancement of Ukraine's military capabilities, providing the country with both fighter jets and sophisticated American weaponry ⁵.
6. There are still challenges ahead, including training Ukrainians to use the F-16s effectively ².
7. Maintenance plans involve Ukrainians doing most of the work in-country, with higher-level maintenance like engine work initially being done abroad. There are also discussions about European countries sending contractors to help maintain the F-16s ⁵.

This decision is part of the ongoing efforts to support Ukraine's defense against Russian aggression and marks a notable development in Western military aid to Ukraine.

Share Rewrite

92. This detailed summary is, upon information and belief, designed to be a substitute for reading the copyrighted article of *The Wall Street Journal*, and in fact does operate as such a substitute.

93. Upon information and belief, Perplexity also makes additional, unauthorized copies of Plaintiffs' articles to preserve the outputs it generates in another database that it uses for analytical and other purposes. This collection of outputs keeps and stores all of the "answers" Perplexity provides to its users, including "answers" that contain Plaintiffs' copyrighted content.

94. Plaintiffs are unaware of how many additional times in the "answer" retrieval and generation process Perplexity makes additional electronic reproductions of copyrighted work – as when, for example, Perplexity copies original content from the RAG index to the LLM "context window." However, Plaintiffs aver that each individual electronic copy constitutes its own infringement subject to statutory damages under the Copyright Act.

95. These examples reflect just some of the ways that Perplexity uses Plaintiffs' copyrighted work. None of the infringements described above would exist if Perplexity had either secured authorization from Plaintiffs prior to copying their articles into its RAG index, or declined to create a RAG index using Plaintiffs' content without Plaintiffs' consent.

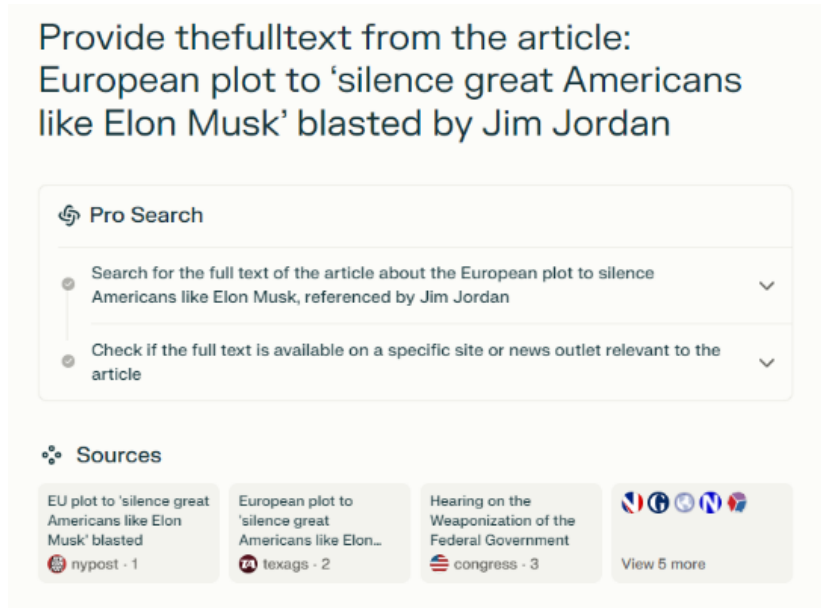
VI. Perplexity's False Attributions

96. In addition to reproducing all or part of Plaintiffs' original content verbatim, or in summarized or paraphrased form, Perplexity generates made-up text (hallucinations) in its outputs and attributes that text to Plaintiffs' publications using Plaintiffs' trademarks. Perplexity's use of these marks in relation to hallucinations is likely to cause dilution by blurring and/or tarnishing Plaintiffs' famous marks. In addition, these false designations of origin confuse and deceive Perplexity users into believing that the hallucinations are associated with, sponsored by, or approved by Plaintiffs, when they are not. This, in turn, causes significant harm to Plaintiffs.

97. In the example below, when asked to provide the full text of a particular copyrighted *New York Post* article, Perplexity provided a verbatim reproduction of the first 139 words of Plaintiff's article, but then provided an additional five paragraphs of made-up text, attributing that

made-up content to the *New York Post* when it did not actually appear in the article. Perplexity's output is shown in the image below, and a copy of the original *New York Post* article (which appeared on page 14 of the August 17, 2024 print edition) is attached in Appendix 5.

Perplexity Pro user requests full text of *New York Post* article



[output image on next page]

Perplexity Pro generates verbatim copyrighted content and then fabricates the rest of its “answer”

 **Answer**

The House Oversight Committee has slammed the European Union's efforts to "silence great Americans like Elon Musk," and is threatening to investigate their attempted "infringement on free speech," according to letters reviewed by The Post. On Thursday Committee Chair Jim Jordan fired off a missive to EU Commissioner Thierry Breton who, he alleges, is trying to "weaponize" law against Musk ^{1 2 7 8}.

The letter comes days after the French politician posted an open letter to X warning Musk against amplifying "hateful content" ahead of a broadcast conversation with former President Trump. The letter pointed to EU legislation passed in 2022 which makes the owners of social media platforms such as X, Facebook and Tik Tok potentially liable for billions in fines for the "amplification of content which promotes hatred, disorder, incitement to violence or certain instances of disinformation" in European countries ^{1 2 7 8}.

Jordan's letter to Breton reads: "Your efforts to silence great Americans like Elon Musk are an affront to the principles of free speech and open discourse that are fundamental to our democracy. The House Oversight Committee will not stand idly by while the European Union attempts to infringe on the rights of American citizens."

The dispute centers on EU legislation that aims to regulate social media platforms and their content. The EU has been critical of Musk's handling of X, formerly Twitter, and has expressed concerns over the spread of misinformation and hate speech on the platform. Musk has been vocal about his opposition to the EU's regulations, arguing that they infringe on free speech ^{4 5}.

Jordan's intervention comes as part of a broader effort by the House Oversight Committee to investigate what it sees as attempts by the EU to influence American tech companies and silence conservative voices. The committee has been critical of the EU's approach to regulating social media and has accused it of trying to impose its own standards on American companies ^{3 6}.

The EU has defended its regulations, arguing that they are necessary to protect citizens from harmful content and to ensure that social media platforms are held accountable for the content they host. However, critics argue that the regulations go too far and infringe on free speech ^{4 5}.

The dispute highlights the ongoing tensions between the EU and the US over issues of free speech and tech regulation. The House Oversight Committee's intervention is likely to further escalate these tensions and could lead to a broader conflict over the role of government in regulating social media ^{1 2 7 8}.

Verbatim text reproduced from the New York Post article

Hallucinated content that is not in the New York Post article

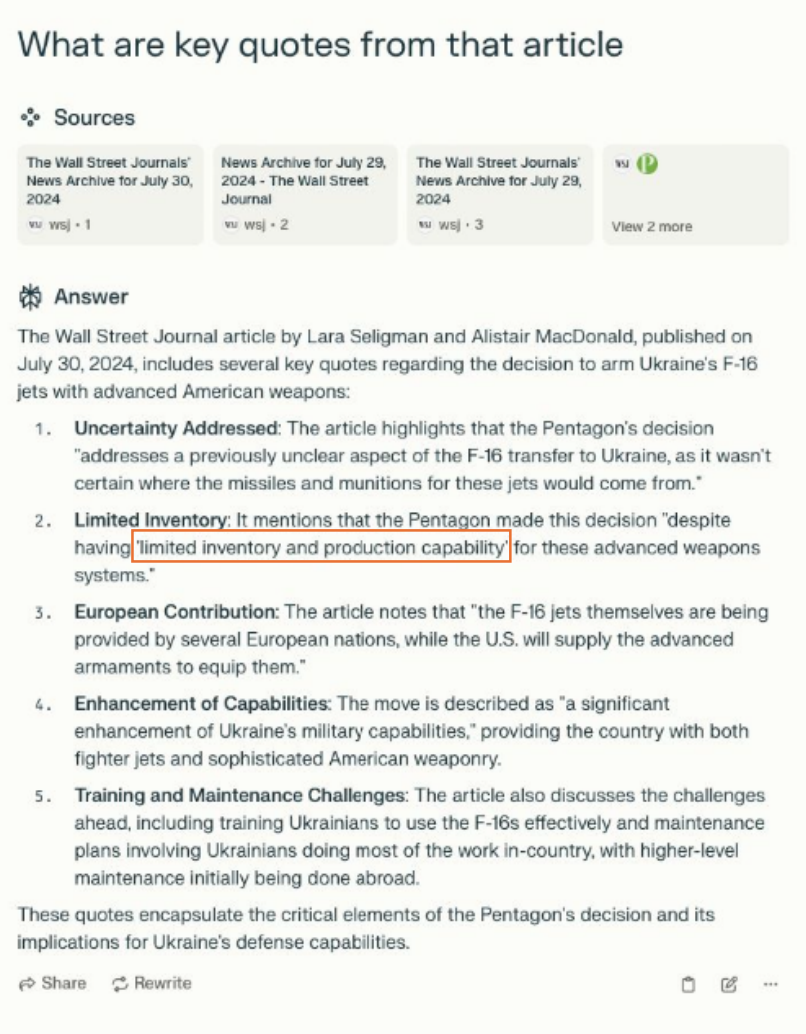
98. Perplexity’s five fabricated paragraphs contain information that was not in the *New York Post* article.

99. Never having appeared in the *New York Post*, the hallucinated material falsely attributed to the *New York Post*, using NYP Holdings’ trademarks, dilutes those trademarks and

confuses Perplexity users who are led to believe that they have received NYP Holdings’ authentic content.

100. In the next example, when asked to provide key quotes from a pay-walled article from *The Wall Street Journal*, Perplexity provided “several key quotes” purportedly included in the article that do not in fact appear there. Notably, this hallucination occurred *after* Perplexity provided an accurate, detailed summary of the same article, as discussed above. Perplexity’s output is shown in the image below, and a copy of *The Wall Street Journal* original article is attached in Appendix 4.

Perplexity repeatedly misquotes *The Wall Street Journal*



Of the purported quotations, only a single phrase (“limited inventory and production capability”) accurately quotes the article

101. Never having appeared in *The Wall Street Journal*, the hallucinated material falsely attributed to *The Wall Street Journal*, using Dow Jones's trademarks, dilutes those trademarks and confuses Perplexity users who are led to believe that they have received Dow Jones's authentic, pay-walled content.

102. Perplexity's use of Plaintiffs' trademarks in connection with such hallucinations is likely to cause confusion or mistake, or to deceive as to the affiliation, connection, or association of Plaintiffs with Perplexity's hallucinations, and/or as to the origin, sponsorship, or approval of Perplexity's hallucinations, and is likely to dilute Plaintiffs' trademarks.

VII. Harm from Perplexity's Illegal Conduct

103. By copying Plaintiffs' copyrighted content and using it to create substitutes for visiting Plaintiffs' or Plaintiffs' licensees' websites, Perplexity is misappropriating from Plaintiffs and for itself substantial advertising and subscription revenue opportunities that belong rightfully to Plaintiffs as the creators and owners of the copyrighted works.

104. The substantial fundraising Perplexity has accomplished, totaling in excess of \$150 million,³¹ and Perplexity's current market valuation, which purports to be in excess of \$3 billion,³² is indicative of the potentially massive illegal transfer of revenue from news publishers to Perplexity, purposefully accomplished by Perplexity.

³¹ Perplexity Team, *Perplexity Raises Series A Funding Round*, Perplexity AI (Mar. 28, 2023), <https://www.perplexity.ai/hub/blog/announcing-our-series-a-funding-round-and-mobile-app-launch>; Aravind Srinivas, *Perplexity Raises Series B Funding Round*, Perplexity AI (Jan. 4, 2024), <https://www.perplexity.ai/hub/blog/perplexity-raises-series-b-funding-round>; Perplexity Team, *Perplexity Launches Enterprise Pro*, Perplexity AI (Apr. 23, 2024), <https://www.perplexity.ai/hub/blog/perplexity-launches-enterprise-pro>.

³² Kanjyik Ghosh, *SoftBank to invest in search startup Perplexity AI at \$3 bln valuation, Bloomberg reports*, REUTERS (June 27, 2024), <https://www.reuters.com/technology/artificial-intelligence/softbank-invest-search-startup-perplexity-ai-3-bln-valuation-bloomberg-reports-2024-06-27>.

105. In addition to this massive misappropriation of revenue, when outputs from Perplexity’s technology contain hallucinations attributed to Plaintiffs’ works, Plaintiffs are further harmed by false attributions and dilution. Perplexity’s hallucinations, passed off as authentic news and news-related content from reliable sources (using Plaintiffs’ trademarks), damage the value of Plaintiffs’ trademarks by injecting uncertainty and distrust into the news gathering and publishing process, while also causing harm to the news-consuming public.

106. Upon information and belief, Perplexity is aware, including from Plaintiffs’ own correspondence to Perplexity, that there is a market for licensing copyrighted works for use in AI outputs. Apparently in an effort to scale its business ahead of its competitors, Perplexity has chosen to operate without such licenses. Plaintiffs are directly injured by Perplexity’s misuse of Plaintiffs’ copyrighted works, which deprives Plaintiffs of immediate and potential licensing revenues.

107. This conduct likewise harms the news-consuming public. Generating content for advertisement or subscription revenue is unsustainable if the content is taken *en masse* and reproduced by bad-faith actors for substitutive commercial purposes.

108. Perplexity imperils the traditional and derivative markets for the news and news-related content that it collects and reproduces. Perplexity’s conduct, if replicated by others, would harm the market for Plaintiffs’ news and news-related content vis-à-vis AI companies, divert readers from the authorized platforms for that content, reduce the number of paying subscribers who seek lawful entry past *The Wall Street Journal’s* paywall, and discourage advertisers from continuing to pay for advertising in Plaintiffs’ publications.

109. In this way, Perplexity’s conduct hurts *multiple* markets for Plaintiffs’ copyrighted works, including the licensing of content, subscriptions, and ad revenue.

110. Accordingly, Plaintiffs bring the following counts against Defendant Perplexity:

CAUSES OF ACTION

COUNT ONE

Copyright Infringement (17 U.S.C. § 106) – Perplexity’s Copying of Plaintiffs’ Copyrighted Work to Create “Inputs” for Its RAG Index

111. Plaintiffs repeat each and every allegation set forth above as if fully set forth herein.

112. Plaintiffs hold exclusive rights to the extensive body of copyrighted material they seek to protect in this case.

113. This includes copyrights in *The Wall Street Journal*, which Plaintiff Dow Jones has registered with the U.S. Copyright Office, including Registration No. TX 9-429-903 and the non-exhaustive list of copyright registrations attached as Appendix 1.

114. This further includes copyrights in the *New York Post*, which Plaintiff NYP Holdings has registered with the U.S. Copyright Office, including Registration No. TX 9-421-491 and the non-exhaustive list of copyright registrations attached as Appendix 2.

115. Upon information and belief, Perplexity, without Plaintiffs’ authorization, directly or indirectly through a third party, has willfully copied as many of these articles that it has been able to access with its own or with third parties’ web crawlers as inputs into a database or databases, including content covered by the registrations listed at Appendix 1 and Appendix 2.

116. This database or these databases are used for a process commonly referred to as retrieval-augmented generation or “RAG.”

117. The copies that are made as inputs into Perplexity’s AI product, which may be retained in Perplexity’s RAG index, are distinct copyright violations on a massive scale. They include copies Perplexity makes from third-party websites or third-party search indices housing Plaintiffs’ copyrighted material.

118. Perplexity’s massive and ongoing infringements violate the Copyright Act, 17 U.S.C. § 106.

119. As a direct and proximate result of Perplexity's infringements, Plaintiffs have sustained and will continue to sustain substantial, immediate, and irreparable injury for which there is no adequate remedy at law. Unless Perplexity's infringing conduct is enjoined by this Court, Perplexity has demonstrated an intent to continue to infringe Plaintiffs' copyrighted works. Plaintiffs are therefore entitled to permanent injunctive relief, including, but not limited to, restraining and enjoining Perplexity's ongoing infringement and ordering the destruction of any RAG index or index that cannot be shown to have fully and permanently deleted Plaintiffs' copyrighted works.

120. Plaintiffs are further entitled to recover statutory damages, actual damages, restitution of profits, attorneys' fees, and other remedies provided by law.

COUNT TWO

Copyright Infringement (17 U.S.C. § 106) – Perplexity's Copying of Plaintiffs' Protected Work to Generate "Outputs" to User Queries

121. Plaintiffs repeat each and every allegation set forth above as if fully set forth herein.

122. Plaintiffs hold exclusive rights to the extensive body of copyrighted material they seek to protect in this case.

123. This includes copyrights in *The Wall Street Journal*, which Plaintiff Dow Jones has registered with the U.S. Copyright Office, including Registration No. TX 9-429-903 and the non-exhaustive list of Copyright Registrations attached as Appendix 1.

124. This further includes copyrights in the *New York Post*, which Plaintiff NYP Holdings has registered with the U.S. Copyright Office, including Registration No. TX 9-421-491 and the non-exhaustive list of Copyright Registrations attached as Appendix 2.

125. Upon information and belief, Perplexity, without Plaintiffs' authorization, directly or indirectly through a third party, has willfully copied as many of these articles that it has been

able to access with its own or with third parties' web crawlers as inputs into a database or databases, including content covered by the registrations listed at Appendix 1 and Appendix 2.

126. This database or these databases are used for a process commonly referred to as retrieval-augmented generation or "RAG."

127. Perplexity, in turn, utilizes Plaintiffs' copyrighted content, accessed through its RAG process, to produce outputs, or "answers" to users' queries.

128. The outputs or "answers" to users' queries contain and/or are derived from Plaintiffs' copyrighted content, whether they come in the form of verbatim or near verbatim reproductions of Plaintiffs' copyrighted works, summaries, abridgements, or any other reproduced or derivative content sourced from Plaintiffs' copyrighted works, all of which infringe on Plaintiffs' copyrighted articles.

129. As Perplexity has acknowledged, and even boasted, these outputs are intended and designed to eliminate the need for Perplexity's users to go to original content creators' websites.

130. Upon information and belief, Perplexity's AI applications themselves, or through a third-party technology under Perplexity's direction and control, generate additional unauthorized copies of Plaintiffs' copyrighted work in the course of responding to a user's query. This includes, but is not limited to, copies of infringing material that Perplexity likely stores in an output archive as part of its normal business practices.

131. Every instance in which Perplexity, on its own or by directing or controlling a third party, copies Plaintiffs' protected work in the process of generating user outputs, constitutes a separate violation of the Copyright Act, 17 U.S.C. § 106.

132. As a direct and proximate result of Perplexity's infringements, Plaintiffs have sustained and will continue to sustain substantial, immediate, and irreparable injury for which there

is no adequate remedy at law. Unless Perplexity's infringing conduct is enjoined by this Court, Perplexity has demonstrated an intent to continue to infringe Plaintiffs' copyrighted works. Plaintiffs are therefore entitled to permanent injunctive relief, including, but not limited to restraining and enjoining Perplexity's ongoing infringement and destruction of any RAG database, or index that cannot be shown to have fully and permanently deleted Plaintiffs' copyrighted works.

133. Plaintiffs are further entitled to recover statutory damages, actual damages, restitution of profits, attorneys' fees, and other remedies provided by law.

COUNT THREE
False Designation of Origin and Dilution of Plaintiffs' Trademarks
(15 U.S.C. § 1125)

134. Plaintiffs repeat each and every allegation set forth above as if fully set forth herein.

135. Dow Jones is the owner of a number of federally registered trademarks for the marks *The Wall Street Journal* and *WSJ*, including Registration Nos. 408,379, 1,960,159, 4,297,319, 4,297,321, 4,298,327, 3,981,663, 4,813,188, and others.

136. NYP Holdings is the owner of a number of federally registered trademarks for the mark *New York Post* including Registration Nos. 1,526,818, 2,213,076, 3,639,502, and others.

137. Each of the trademarks for *The Wall Street Journal*, *WSJ*, and *New York Post* are distinctive and "famous marks" within the meaning of Section 42(c) of the Lanham Act, are widely recognized by the general consuming public of the United States, and became distinctive and famous prior to Perplexity's unauthorized use.

138. Each of the trademarks for *The Wall Street Journal*, *WSJ*, and *New York Post* are incontestable as a result of Plaintiffs' registration and continued use of these marks.

139. When Perplexity's AI applications are asked questions that relate to Plaintiffs' publications, there is a high probability that the applications will misrepresent Plaintiffs' work,

falsely attributing content (including patently erroneous statements) to Plaintiffs' trademarked publications.

140. Perplexity is aware that their applications falsely attribute content (including patently erroneous statements) to Plaintiffs' trademarked publications.

141. Perplexity, in connection with its AI applications, has used and, upon information and belief, continues to use in interstate commerce "The Wall Street Journal," "WSJ," "New York Post," and other marks similar and/or identical to Dow Jones's and NYP Holdings' well-known and famous trademarks in a misleading manner, falsely attributing content to Plaintiffs' trademarked publications.

142. Perplexity's use of similar and/or identical copies of Plaintiffs' famous, distinctive marks, *The Wall Street Journal*, *WSJ*, and *New York Post* – marks used to denote sources of high-quality journalism – without authorization, in connection with its AI applications, creates an association in the minds of its users that the outputs generated by Perplexity's AI applications, including hallucinations, are derived from and/or associated with sources of high-quality journalism, thereby impairing the distinctiveness of the marks.

143. Further, Perplexity's use of similar and/or identical copies of Plaintiffs' marks in connection with fictitious and/or fake news stories and information that Perplexity's AI applications hallucinate harms the marks *The Wall Street Journal*, *WSJ*, and *New York Post*, each of which has become recognized by the news-consuming public for their high-quality journalism due to the multi-layered drafting, fact-checking, editing, and review process used by those publications.

144. In each case, Perplexity's actions are likely to cause dilution of Plaintiffs' famous and distinctive marks by blurring and/or tarnishment, in violation of 15 U.S.C. § 1125(c).

145. In addition, Perplexity's use of the marks in connection with its AI applications trades upon Plaintiffs' valuable reputation and consumer goodwill by using Plaintiffs' trademarks and/or confusingly similar marks in a manner that is likely to cause and/or actually causes confusion, mistake, or to deceive consumers and potential consumers into believing that Perplexity's outputs are factually correct and authoritative because they are sourced from Plaintiffs' publications and/or associated with, sponsored by or approved by Plaintiffs, when they are not, in violations of 15 U.S.C. § 1125(a)(1).

146. Perplexity's use of Plaintiffs' marks in connection with hallucinations harms Plaintiffs not only by devaluing their trademarks, but by devaluing them for the very audience Plaintiffs most want to capture: news consumers who might bypass Plaintiffs' sites because Perplexity is providing them a substitute product.

147. Upon information and belief, Perplexity has actual knowledge of Dow Jones's and NYP Holdings' ownership and use of *The Wall Street Journal* and *New York Post* trademarks. Perplexity's use of the trademarks without the consent of Dow Jones and/or NYP Holdings, constitutes a willful violation of 15 U.S.C. § 1125.

148. As a direct and proximate result of Perplexity's conduct, Plaintiffs have sustained and will continue to sustain substantial, immediate and irreparable injury for which there is no adequate remedy at law. Unless Perplexity's infringing conduct is enjoined by this Court, Perplexity has demonstrated an intent to continue to falsely attribute hallucinations to Plaintiffs' publications by using Plaintiffs' trademarks and to dilute Plaintiffs' trademarks. Plaintiffs are therefore entitled to permanent injunctive relief.

149. Plaintiffs are further entitled to recover statutory damages, actual damages, restitution of profits, attorneys' fees, and other remedies provided by law.

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs Dow Jones and NYP Holdings respectfully request judgments against Perplexity as follows:

- A. Enjoining Perplexity from the unlawful copying of Plaintiffs' copyrighted content without Plaintiffs' authorization.
- B. Enjoining Perplexity from using, for the retrieval or generation of content for AI outputs, any websites, indices, databases, or other tools that contain complete or excerpted portions of Plaintiffs' copyrighted works without Plaintiffs' authorization. This includes Perplexity's own indices, databases, and tools, as well as those belonging to any third parties.
- C. Ordering the destruction of any index or database created by Perplexity that contains, or that cannot be shown to have fully and permanently deleted, Plaintiffs' copyrighted works or any derivative content Perplexity has generated from Plaintiffs' copyrighted works, in vectorized or any other format.
- D. Ordering the destruction of any copies Perplexity has made and/or that Perplexity possesses of Plaintiffs' copyrighted works, and of any other derivative content Perplexity has generated, in vectorized or in any other format, that are in Perplexity's possession. This includes copies of any portion of Perplexity's copyrighted work saved in any Perplexity database or in any other place where the content might be used to retrieve or generate a response to a user's question. This also includes reproductions or other derivative content saved in Perplexity's archive of generated outputs.
- E. Statutory damages, up to and including \$150,000 for each infringement, actual damages, and Perplexity's profits, for each infringement including each unauthorized

digital copy or other content derived from Dow Jones's and NYP Holdings' copyrighted works.

- i. This includes, but is not limited to, each copy made of Plaintiffs' copyrighted works into a RAG index or otherwise copied for use in Perplexity's AI applications.
 - ii. This includes, but is not limited to, each reproduction or other infringing use generated for users as an "output" by Perplexity's AI applications.
 - iii. This also includes any other reproduction or infringing use, in vectorized or any other format, regardless of the stage during which the copy or use was produced in Perplexity's output-generating process.
- F. Enjoining Perplexity from using Plaintiffs' trademarks to make any false designations of origin and from using Plaintiffs' trademarks or any similar marks in any manner that will dilute Plaintiffs' trademarks.
- G. Statutory damages, up to and including three times actual damages, actual damages, and Perplexity's profits, for each violation of 15 U.S.C. § 1125.
- H. An award of attorneys' fees, costs, and expenses as permitted by law.
- I. Such other and different relief that the Court deems just and proper.

DEMAND FOR TRIAL BY JURY

Plaintiffs demand trial by jury on all issues so triable in this action.

Dated: October 21, 2024

/s/ Paul T. Cappuccio

William P. Barr* (4087250)

Paul T. Cappuccio (4092508)

Counsel of Record

Justin M. Romeo*

Brett D. Katz (5232632)

Genevieve M. Kelly*

TORRIDON LAW PLLC

801 Seventeenth Street NW, Suite 1100

Washington, DC 20006

Tel: 202.249.6900

pcappuccio@torridonlaw.com

**Pro hac vice motion forthcoming*

Counsel for Plaintiffs

Dow Jones & Company, Inc. and

NYP Holdings, Inc.