**UNITED STATES DISTRICT COURT**
**SOUTHERN DISTRICT OF NEW YORK**

| | |
|---|---|
| ENCYCLOPAEDIA BRITANNICA, INC., and MERRIAM-WEBSTER, INC.,<br><br>        Plaintiffs,<br><br>   v.<br><br>PERPLEXITY AI, INC.,<br><br>        Defendant. | No. 1:25-cv-07546-JLR |

<u>**PLAINTIFFS' MEMORANDUM OF LAW IN OPPOSITION TO PERPLEXITY'S MOTION TO DISMISS**</u>

**TABLE OF CONTENTS**

# TABLE OF AUTHORITIES

**Page(s)**

## Cases

iv

**Statutes**

**Rules**

I.      **INTRODUCTION**

This case lies at the intersection of emerging generative artificial intelligence technology and copyright law. Plaintiffs Encyclopaedia Britannica, Inc. ("Britannica") and Merriam-Webster, Inc. ("Merriam-Webster") (together, "Plaintiffs") allege claims of copyright infringement against Perplexity AI, Inc. ("Perplexity"), the developer of a so-called "answer engine" that threatens to replace online publishers like Plaintiffs using unauthorized derivatives of their own works. Perplexity does so by engaging in massive copying of Plaintiffs' content, providing AI-generated summaries of Plaintiffs' content to users who search for such information, and thus disincentivizing users to visit Plaintiffs' websites to explore information responsive to their queries. By cannibalizing traffic to Plaintiffs' websites in this manner, Perplexity also starves Plaintiffs of revenue that depends on web traffic, *i.e.* from selling advertisements on their websites and subscriptions to their content. Specifically, Plaintiffs allege three claims against Perplexity, one for the use of their copyrighted materials to create "inputs" for Perplexity's retrieval-augmented generation ("RAG") content, a second for the use of their copyrighted materials to create "outputs" to user queries, and a third for trademark infringement. Dkt. 1 ("Compl.") at ¶ 9. In its partial motion to dismiss, Perplexity challenges only the second claim, the copyright "output" claim.

Perplexity seeks to dismiss this claim by inaccurately depicting its answer engine as an "automated" or passive computer program that merely executes user queries. By Perplexity's telling, it does not exercise sufficient volitional conduct when its answer engine spits out answers that are verbatim or near-verbatim copies of Plaintiffs' copyrighted works. Perplexity argues that its user, not Perplexity itself, exercises the volitional conduct required to find direct infringement. This tactic of shifting liability to its users is meritless for at least three reasons.

First, Perplexity's attempt to divorce the "inputs" into its program from the "outputs" of that program contradicts Plaintiffs' detailed allegations. Although Perplexity seeks to portray its technology in discrete, piecemeal parts, the Complaint makes clear that copying of "inputs" by Perplexity and generation of "outputs" in response to user queries are part of an integrated system that operates as a whole to constitute the infringing answer engine. Tellingly, Perplexity does not even challenge the copyright "input" claim, which also favors denying its motion to dismiss the "output" claim.

Second, Perplexity relies on inapposite caselaw to argue that it lacks volitional conduct. While Perplexity correctly recognizes that volitional conduct is required to find direct liability for copyright infringement, it overlooks that its own conduct—in scraping and crawling Plaintiffs' content to curate the universe of copyrighted material from which to generate output— demonstrates sufficient volitional conduct under well-established authority. *See, e.g.*, *EMI Christian Music Grp., Inc. v. MP3tunes, LLC*, 844 F.3d 79 (2d Cir. 2016); *Capitol Records, LLC v. ReDigi Inc.*, 934 F. Supp. 2d 640 (S.D.N.Y. 2013), *aff'd*, 910 F.3d 649 (2d Cir. 2018).

Third, Perplexity's argument about insufficient volitional conduct focuses entirely on the subset of user queries that specifically request Plaintiffs' content (*e.g.*, a user asking, "How does Merriam Webster define plagiarize?"). Perplexity ignores Plaintiffs' allegation that "[e]ven when the user does not prompt Perplexity to limit its answers to Britannica's content, Perplexity often does so." Compl. at ¶ 77. Even if Perplexity is correct that it lacks volitional conduct when it generates unauthorized copies of Plaintiffs' content in response to user queries that specifically request that content (it is not), Perplexity has no argument for its copying when the user does not prompt the answer engine for Britannica or Merriam-Webster material.

Perplexity also makes a last-ditch attempt to dismiss the copyright "output" claim on the ground that the Complaint does not plead every single example of every single instance of infringement. This supposed lack of examples, according to Perplexity, leaves the Court unable to examine the "substantial similarity" between Plaintiffs' content and Perplexity's outputs. This argument is both factually and legally deficient. Factually, it overlooks Plaintiffs' allegation that "Perplexity has copied hundreds of thousands of Plaintiffs' copyrighted articles" and even potentially "all of Plaintiffs' copyrighted content" since Perplexity's founding in 2022. *Id*. at ¶¶ 52, 65. It would be infeasible for Plaintiffs to provide every single example of copyright infringement given the sheer scale of the unauthorized copying and the fact that records of infringing outputs are solely in Perplexity's control. Legally, there is no pleading standard that requires a copyright owner to allege every instance of every infringement to survive dismissal. Likewise, Perplexity's motion misconstrues the doctrine of substantial similarity, a fact-bound question commonly reserved for the jury except in narrow circumstances where "the works in question are attached to a plaintiff's complaint," and "the district court [can] consider the similarity between those works in connection with a motion to dismiss, because the court has before it all that is necessary in order to make such an evaluation." *Peter F. Gaito Architecture, LLC v. Simone Dev. Corp.*, 602 F.3d 57, 64 (2d Cir. 2010).

Plaintiffs respectfully request that the Court deny Perplexity's Motion to Dismiss Count II of the Complaint. Dkt. 30 ("Mot.").

## II.    BACKGROUND

Plaintiffs have published the world's leading English-language encyclopedia and dictionary for well over a hundred years. Compl. at ¶¶ 34, 37. Today, Plaintiffs offer this information as digital content. *Id*. at ¶¶ 2–5, 34–38. Plaintiffs employ and rely on hundreds of

human employees, including writers, editors, researchers, and content creators, to generate their original digital content. *Id*. at ¶ 39. In order to protect the immense resources Plaintiffs have invested to create and continuously update that content, Plaintiffs have registered multiple copyrights and trademarks, including copyrights in nearly 100,000 online articles. *Id*. at ¶¶ 40–43.

Perplexity is a generative artificial intelligence company that offers a product it calls an "answer engine." Compl. at ¶ 1. As alleged in the Complaint, Perplexity threatens Plaintiffs' business by "free rid[ing]" on the enormous investment that Plaintiffs make to create and maintain their content "by cannibalizing traffic to Plaintiffs' websites with AI-generated summaries of Plaintiffs' own content." *Id.* at ¶ 6. Traditional search engines serve as intermediaries that connect people browsing the internet with online sources that provide relevant information. *Id*. Perplexity's answer engine is different. Instead of connecting people with online sources, Perplexity copies those original sources at a massive scale, feeds the sources to Perplexity's retrieval-augmented generation ("RAG") model, and then "repackages the original content in written responses to users." *Id*. at ¶ 8.

The bottom layers of Perplexity's answer engine are large language models ("LLMs"), including LLMs created by OpenAI, Anthropic, and Perplexity itself. *Id*. at ¶¶ 46–48. LLMs are trained on vast quantities of data, and once trained can generate natural-language text based on the knowledge foundation the LLMs have from the universe of materials they were trained on. *Id*. at ¶¶ 47, 49–50. After it is trained, however, an LLM does not have access to dynamic external information, such as up-to-date content. *Id*. at ¶ 50. To bridge that gap, Perplexity's answer engine uses the RAG process in connection with LLMs to incorporate up-to-date external sources of online information, such as copied material from Plaintiffs' websites. *Id*. at ¶¶ 49–50. Plaintiffs allege that in its quest for market dominance, Perplexity crawls and scrapes with impunity,

4

notwithstanding public outcry over its "surreptitious[]" and "unethical" practice of "using stealth, undeclared crawlers to evade website no-crawl directives." *Id.* at ¶¶ 58–70 & nn. 38, 49, 70.

As alleged, Perplexity infringes Plaintiffs' copyrighted materials, including by scraping Plaintiffs' websites in direct violation of Plaintiffs' terms of use, intentionally copying hundreds of thousands of Plaintiffs' copyrighted works (and potentially "all of Plaintiffs' copyrighted content" since 2022), and using those copied works both as inputs for its RAG content and as outputs delivered in response to user queries. *See, e.g.*, Compl. at ¶¶ 52, 65, 70, 112–15.

The RAG process "is not possible without high-quality, fact-checked content like that of Plaintiffs." *Id.* at ¶ 57. "Perplexity chooses only high-quality sources," such as those of Plaintiffs, "as RAG Content to 'compile[] the most relevant insights into a coherent, easy-to-understand answer.'" *Id.* at ¶ 50; *see also id.* at ¶ 64 (alleging how Perplexity boasts about the "credible sources" that it copies, including "citations from reputable news organizations, academic publications, and established content sources"). Perplexity distinguishes itself from other generative AI software on the market by emphasizing its RAG, or "grounding," functionality that utilizes Plaintiffs' copyrighted content. *Id.* at ¶¶ 53–57.

Contrary to Perplexity's motion which depicts the "input" and "output" ends of its infringing answer engine as discrete parts, Plaintiffs' allegations reflect that the two cannot be divorced. They together constitute the infringing "answer engine." When a user asks a question to Perplexity's answer engine, it "obtain[s] and cop[ies] content from its search index relating to the prompt" and "combine[s] the original prompt with the retrieved copied content in order to provide additional context." *Id.* at ¶ 49. The answer engine provides that combined data, which includes Plaintiffs' copied materials, to an LLM that generates a natural-language response for the user. *Id.* In fact, in Count II itself, Plaintiffs allege that Perplexity (1) "has willfully copied as many of

Plaintiffs' articles and content that it has been able to access with its own or with third parties' web crawlers as inputs into database(s) or index(es)," (2) utilizes those databases or indexes for its RAG process, and (3) "*uses Plaintiffs' copyrighted content, accessed through its RAG process, to produce outputs*, or 'answers' to user queries." *Id*. at ¶¶ 112–14 (emphasis added). The "input" and "output" mechanisms cannot be disentangled because, when a user makes a request, the "outputs" that copy Plaintiffs' materials are generated using a RAG process that incorporates the copied material that Perplexity scraped from Plaintiffs' websites.

## III.   LEGAL STANDARD

When considering a Fed. R. Civ. P. 12(b)(6) motion to dismiss, the court must "accept as true all factual allegations set forth in the complaint" and "draw all reasonable inferences in favor of" the plaintiff. *Cnty. of Suffolk, New York v. First Am. Real Estate Sols*., 261 F.3d 179, 184 (2d Cir. 2001). The complaint must "state a claim to relief that is plausible on its face," meaning that the "plaintiff pleads factual content that allows the court to draw the reasonable inference that the defendant is liable for the misconduct alleged." *Ashcroft v. Iqbal,* 556 U.S. 662, 678 (2009).

This court "has been strictly proscribed" from "subject[ing] copyright plaintiffs to a heightened level of pleading." *Warren v. John Wiley & Sons, Inc*., 952 F. Supp. 2d 610, 617 (S.D.N.Y. 2013); *see also Energy Intelligence Group, Inc. v. Jefferies, LLC*, 101 F. Supp. 3d 332, 340 (S.D.N.Y. 2015) ("[C]opyright claims are not subject to a heightened pleading standard."). "To survive dismissal, the complaint must allege: (1) which specific original works are the subject of the copyright claim, (2) that plaintiff owns the copyrights in those works, (3) that the copyrights have been registered in accordance with the statute, and (4) by what acts and during what time the defendant infringed the copyright." *Energy Intelligence,* 101 F. Supp at 338 (internal citations omitted).

IV.    **ARGUMENT**

Plaintiffs state a claim for direct infringement based on Perplexity's volitional conduct through its creation and use of a RAG process that relies on Plaintiffs' content to generate outputs. Perplexity seeks to dismiss Plaintiffs' well-pled copyright output claim based on an inapplicable body of caselaw involving passive platforms or automated systems that bear no resemblance to Perplexity's answer engine. Indeed, Perplexity ignores that its answer engine can deliver output showing Britannica's content precisely because it has independently crawled and scraped such content regardless of the user query. Through its own volitional conduct, Perplexity curates the universe of content from which to generate output for the user.

In any event, Perplexity's motion overlooks Plaintiffs' allegation that "[e]ven when the user does not prompt Perplexity to limit its answers to Britannica's content, Perplexity often does so." Compl. at ¶ 77. Both reasons independently defeat Perplexity's motion.

A.    **Plaintiffs Sufficiently Allege that Perplexity's Infringement Results from Perplexity's Volitional Conduct.**

1.    *MP3tunes*, not *Cablevision*, provides the relevant framework to assess Perplexity's direct infringement.

Perplexity's motion is fundamentally misguided because it relies on an irrelevant body of caselaw to argue that its intentional copying and use of Plaintiffs' copyrighted works does not constitute volitional conduct. As Perplexity recognizes, "[w]hen there is a dispute as to the author of an allegedly infringing instance of reproduction, [the court] direct[s] [its] attention to the volitional conduct that causes the copy to be made." *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 131 (2d Cir. 2008) ("*Cablevision*") (Mot. at 8–10). In that case, Cablevision, the cable provider, offered a digital video recorder ("DVR") system that allowed its users to record programs broadcast by individual television channels. Various owners of copyrighted programs

7

brought suit alleging that Cablevision directly infringed their copyrights by offering the DVR system. The court ruled that Cablevision lacked sufficient volitional conduct for direct liability because the user, not Cablevision, selected which precise programs to record. Cablevision's "control is limited to the channels of programming available to a customer and not to the programs themselves. Cablevision has no control over what programs are made available on individual channels or when those programs will air, if at all." *Id.* at 132. Perplexity, by contrast, exercises precisely the control that was lacking in *Cablevision* by choosing the content that is available for users to request. While *Cablevision* stands for the unremarkable proposition that, as Perplexity frames it, "the law distinguishes between services that automatically make unlawful copies ***at the user's request*** and systems that do the same ***without a user request***," Mot. at 8 (emphasis in original), it does not absolve the provider of the former type of service who selects the unlawful copies that may be requested by the user.

The Complaint makes clear that Perplexity's answer engine supplies unauthorized copies of Plaintiffs' works for users to request prior to any user having done so. Even before any user submits a prompt to Perplexity's answer engine, Perplexity has already sent out crawlers (and often "undeclared" and "stealth" crawlers) to Plaintiffs' websites to scrape Plaintiffs' copyrighted content for its answer engine. Compl. at ¶¶ 59, 66; *see also id.* at ¶¶ 67–70. Only upon that foundation does a user query "kick[] off" the process where "Perplexity's AI products (1) 'obtain and copy content from its search index relating to the prompt;' (2) 'combine the original prompt with the retrieved copied content in order to provide additional context;' and (3) 'provide the combined data to an LLM, which generates a natural-language response.'" Mot. at 5 (quoting Compl. ¶ 49). There would be nothing for the user to query without Perplexity's selection of infringing content in the first place and its decision to use that infringing content in its RAG system.

8

This circuit, like other courts nationwide, has not hesitated to recognize volitional conduct in such circumstances. For instance, in *Capitol Recs., Inc. v. MP3tunes, LLC*, 48 F. Supp. 3d 703 (S.D.N.Y. 2014), this court upheld a jury's finding that MP3tunes—an integrated music service that allowed users to store, stream, or download music from online storage "lockers"—was liable for direct infringement of plaintiffs' copyrights in album cover art. *Id.* at 711, 720. The Second Circuit affirmed. *See EMI Christian Music Grp., Inc. v. MP3tunes, LLC.*, 844 F.3d 79 (2d Cir. 2016). Both courts rejected MP3tunes' argument that it lacked volition "because the retrieval of that art was not directed by MP3tunes but rather obtained from Amazon.com at the direction of the user when the user selected a song." *Id.* at 96. Rather, the courts recognized that the software "was designed to retrieve one aspect of a copyrighted work (the album art) whenever a user uploaded another aspect of a copyrighted work (the song)." *Id.* "In other words, the system retrieved a copyrighted item that a user did not request, frequently without the user's knowledge of the copyrighted nature of the item. This constituted enough evidence, in our view, that copying of the cover art was directed by MP3tunes, not users."[1] *Id.* Perplexity devotes one parenthetical to *EMI Christian*, noting only that the court "f[ound] volitional conduct where [the] system retrieved copyrighted material ***without*** a user request." Mot. at 8. This simplistic description omits that the user exercised some volition in selecting a song, based on which MP3tunes retrieved the cover art associated with the song. Notwithstanding this level of user volition built into the MP3tunes software, the Second Circuit found that the "copying of the cover art was directed by MP3tunes, not users." 844 F.3d at 96.

---

[1] Any attempt to distinguish *MP3tunes* on the ground that Perplexity's users did have "knowledge of the copyrighted nature" of the answers generated by Perplexity is not permissible at the 12(b)(6) stage, where all factual disputes must be resolved in the nonmovants' favor. *See Cnty. of Suffolk,* 261 F.3d at 184.

9

Perplexity's answer engine is more akin to the MP3tunes software than Cablevision's DVR system. Even where a user specifies Britannica content as the source for a requested answer in a prompt, it is Perplexity's answer engine that chooses which Britannica article or articles are responsive to the query and which portions of them to summarize in its answers. The Complaint, for example, describes Perplexity selecting a specific article and choosing which passages to excerpt in response to more general requests to provide information on a topic using "Britannica content" and provide unspecified "exact passages" from the source that Perplexity chose. *See* Compl. ¶ 75. Unlike Cablevision, which exercised no control over the programs that users chose to record, Perplexity "curates a hand-picked selection of infringing content for users to [query]." Katherine Lee *et al.*, "Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain," *Journal of the Copyright Society of the U.S.A.*, 72 J. Copyright Soc'y 251 (2025), at 349 ("Lee *et al.*").

*MP3tunes* is not the only case to conclude that some software system exercises sufficient volitional conduct for direct liability. In *Capitol Records, LLC v. ReDigi Inc.*, 934 F. Supp. 2d 640 (S.D.N.Y. 2013), *aff'd*, 910 F.3d 649 (2d Cir. 2018), this court concluded that an "online marketplace for digital used music" engaged in volitional conduct due to its selection of copyrighted materials and active delivery of those materials to the users. *Id.* at 645, 657. The platform's "founders programmed their software to *choose* copyrighted content," "provided the infrastructure for its users' infringing sales," and "affirmatively brokered sales by connecting users who are seeking unavailable songs with potential sellers." *Id*. at 657 (emphasis added). The fact that the scanning process was "automated" was "a distinction without a difference." *Id*. Just like the online platform in *ReDigi*, Perplexity chooses to scrape Plaintiffs' copyrighted content and copy that content for use in the RAG process; "provide[s] the infrastructure" for users to receive

10

infringing content by selecting that content for its RAG process in the first place; and "affirmatively" generates infringing responses to users' requests. As this court concluded about the *ReDigi* platform, Perplexity's "fundamental and deliberate" role has "transformed it from a passive provider of a space in which infringing activities happened to occur to an active participant in the process of copyright infringement." *Id*. (quoting *Arista Records LLC v. Usenet.com, Inc*., 633 F. Supp. 2d 124, 148 (S.D.N.Y. 2009)) (cleaned up).

This court and the Second Circuit have continued to emphasize that software systems engage in volitional conduct when the platform has an active, intentional role in curating copyrighted information or providing the infrastructure for its delivery. For example, in *Fox News Network, LLC v. TVeyes, Inc*., 883 F.3d 169 (2d Cir. 2018), the platform engaged in "clear" volitional conduct when it "decide[d] what audiovisual content to record" and then "continuously record[ed] vast quantities of programming, compil[ed] the recorded broadcasts into a database that is text-searchable . . . and allow[ed] its clients to search for and watch (up to) ten-minute video clips that mention terms of interest to the clients."[2] *Id.* at 173, 181. Perplexity's system shares the key characteristics of the platform in *TVEyes*: Perplexity "decides" to scrape Plaintiffs' websites to obtain Plaintiffs' copyrighted materials, incorporates that content into databases or indexes, and uses a RAG process that relies on that copied content to produce outputs in response to specific user requests. *See also BWP Media USA Inc. v. Polyvore, Inc*., 922 F.3d 42, 50 (2d Cir. 2019) (Walker, J., concurring) (cited in Mot. at 9) (an internet service provider "acts volitionally when it

---

[2] This is consistent with other cases that distinguish between basic recording systems like the system in *Cablevision* and more modern streaming and recording platforms. Perplexity cites the dissent from *Am. Broad. Companies, Inc. v. Aereo, Inc*., Mot. at 8, but the Supreme Court held in that case that a web-based streaming platform *did* directly infringe the "transmit clause" of the Copyright Act when it sold "its subscribers a *technologically complex* service that allows them to watch television programs over the Internet at about the same time as the programs are broadcast over the air." 573 U.S. 431, 436 (2014) (emphasis added).

creates a program designed to infringe copyrighted material and selects the copyrighted material that it copies"); *Arista Records*, 633 F. Supp. 2d at 148 (web platform engaged in volitional conduct when "[d]efendants were well aware that digital music files were among the most popular articles on their service, and took active measures to create servers dedicated to mp3 files and to increase the retention times of newsgroups containing digital music files"). Under *MP3tunes* and its progeny, the Complaint sufficiently alleges Perplexity's volitional conduct.

Perplexity seeks to rely on cases involving "automated systems" or passive platforms that strictly execute users' requests. Perplexity cannot avail itself of this line of cases because its answer engine is neither an automated system nor a passive platform. Rather, Perplexity deliberately copies Plaintiffs' works, curates that copied content, and responds to user queries using a RAG process that uses the copied content in ways that exceed the user's explicit control. *See, e.g.*, *White v. DistroKid, LLC*, 766 F. Supp. 3d 451, 462 (S.D.N.Y. 2025) (cited in Mot. at 8) (music distribution website was comparable to a "copy shop" when it enabled users to upload music files and populate them to websites chosen by the user); *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 550–51 (4th Cir. 2004) (cited in Mot. at 8) (internet service providers act as "traditional copying machine[s]" when they are "transmission facilit[ies]" that serve as "conduits from or to would-be copiers and have no interest in the copy itself"); *Bus. Casual Holdings, LLC v. YouTube, LLC*, 2022 WL 837596, at *4 (S.D.N.Y. Mar. 21, 2022) (cited in Mot. at 11–12) (YouTube did not engage in volitional conduct when a third party posted videos to YouTube that copied the plaintiff's own uploaded videos); *Wolk v. Kodak Imaging Network, Inc.*, 840 F. Supp. 2d 724, 742 (S.D.N.Y. 2012) (cited in Mot. at 12), *aff'd*, 569 Fed. App'x 51 (2d Cir. 2014) (photo printing website did not engage in volitional conduct when it displayed an artist's photo transmitted from another website).

The same is true for the out-of-circuit cases that Perplexity relies on, each of which involve platforms passively hosting content that was uploaded by third parties. *See, e.g.*, *Perfect 10, Inc. v. Giganews, Inc.*, 847 F.3d 657, 669 (9th Cir. 2017) (cited in Mot. at 8) (owner of servers that hosted user-uploaded content did not engage in volitional conduct when it did not "play[] any sort of active role in causing the distribution"); *Long v. Dorset*, 854 Fed. App'x 861, 863 (9th Cir. 2021) (cited in Mot. at 9) (Facebook did not engage in volitional conduct by failing to immediately remove infringing posts); *Hunley v. Instagram, LLC*, 73 F.4th 1060, 1062 (9th Cir. 2023) (cited in Mot. at 8) (Instagram was not secondarily liable when third-party websites embedded users' Instagram posts); *Stross v. Twitter, Inc.*, 2022 WL 1843142, at *3 (C.D. Cal. Feb. 28, 2022) (cited in Mot. at 9) (photographer did not state a direct infringement claim against Twitter when photographer did not allege any "nexus" to Twitter's conduct).[3]

Perplexity's answer engine is not akin to the passive, automated platforms in the cases it relies on because, in response to user queries, the model and RAG process make active choices about the sources from which to pull content. The fact that it is not a passive, automated system is demonstrated by Plaintiffs' allegations that Perplexity "mysteriously" omits large swathes of copied content that would be responsive to the query. *See, e.g.*, Compl. at ¶¶ 75–76. A true automated system, like a photocopier, would copy the articles verbatim without leaving out certain sentences or paragraphs from, or adding hallucinations to, the copy.

---

[3] The remaining cases Perplexity relies on are even further afield. *See Abdin v. CBS Broad. Inc.*, 971 F.3d 57, 73 (2d Cir. 2020) (cited in Mot. at 7) (holding that a Star Trek television series is not substantially similar to plaintiff's videogame); *ABKCO Music, Inc. v. Sagan*, 50 F.4th 309, 322 (2d Cir. 2022) (cited in Mot. at 7–8) (holding that district court wrongly applied vicarious liability standard to a direct liability issue).

*VHT, Inc. v. Zillow Group, Inc.*, 918 F.3d 723 (9th Cir. 2019) (cited in Mot. at 8), reinforces why the Complaint here alleges sufficient volitional conduct on Perplexity's part. In *Zillow*, the Ninth Circuit observed that "to demonstrate volitional conduct, a party . . . must provide some 'evidence showing the alleged infringer exercised control (other than by general operation of its website); [or] selected any material for upload, download, transmission, or storage . . . .'" *Id.* at 732 (alterations omitted). Based on this framework, the court found that Zillow did not engage in volitional conduct as to a set of photos that it did not "select" or "exercise control" over, beyond the "general operation of its website." *Id.* at 733 (alterations omitted). By contrast, the court found that Zillow had engaged in volitional conduct as to another set of photos that its own moderators had "selected and tagged" for display. *Id.* at 735. Here, Perplexity crawled, scraped, and selected Plaintiffs' copyrighted content to be used as output for its answer engine. Perplexity thus exercised control over the database of copyrighted content from which to generate output in response to user queries. The Court should not allow Perplexity to shift blame to the end user to avoid liability for its direct infringement of Plaintiffs' copyrights.

At best, even drawing factual inferences in *Perplexity's* favor (which of course is not the appropriate legal standard here), both Perplexity *and* the user may exercise volitional conduct. As contemplated by Lee *et al.* in their article on the intersection of generative AI and copyright law, where the user inputs a query like "frozen 3 screenplay" into an AI tool "that has been trained on screenplays of thousands of films from popular franchises and fine-tuned to optimize its ability to write sequels," the output "will be an infringing derivative work of *Frozen* and *Frozen 2*." Lee *et al.*, *supra*, at 354. In such a case, "*both* the user and provider might be treated as direct infringers." *Id.* The AI model has necessary volition because it "was trained specifically to generate screenplays that incorporate expression from popular franchises. On this view, the service is like

14

a very large archive of copyrighted works . . . ." *Id.* Even on this view, Perplexity cannot absolve itself of direct liability when the user also exercises volition because *both* the user *and* Perplexity are direct infringers. *See Arista Records*, 633 F. Supp. 2d at 158  ("It is well settled in this Circuit that '[a]ll persons and corporations who participate in, exercise control over, or benefit from the infringement are jointly and severally liable as copyright infringers.'" (quoting *Sygma Photo News, Inc. v. High Society Magazine, Inc.*, 778 F.2d 89, 92 (2d Cir. 1985))).

2. <u>Plaintiffs allege that Perplexity generates infringing output even when a user *does not* specifically request Plaintiffs' content.</u>

Even if user queries that specifically request Britannica content somehow absolve Perplexity of direct liability (they do not), Perplexity ignores Plaintiffs' allegation that "[e]ven when the user does not prompt Perplexity to limit its answers to Britannica's content, Perplexity often does so."[4] Compl. at ¶ 77. Plaintiffs further included a specific example of a user asking Perplexity "what are the top 9 mysterious disappearances of people other than Amelia Earhart" and Perplexity regurgitating an identical reproduction of the selection and ordering of such people as in Britannica's article. *See id.* at ¶ 77 and accompanying figures.

This allegation undermines the very premise of Perplexity's motion: that there is a clear distinction between "automatic" systems that generate copies when asked by a user, and systems that do so without a precise request. *See* Mot. at 8. Plaintiffs allege a specific example of Perplexity's answer engine copying Plaintiffs' materials without a user specifically requesting Britannica or Merriam-Webster material. Compl. at ¶ 77. This allegation alone defeats Perplexity's motion even under its own framing of the legal issue. Plaintiffs further allege: "These are merely a few examples of Perplexity generating verbatim or near-verbatim text and identical curation of

---

[4] As discussed in the subsequent section, Plaintiffs need not allege specific examples of each instance of infringement in order to state a claim.

content (including in selection and prioritization of lists) from Britannica's articles. Upon information and belief, there are many others." *Id*. at ¶ 78. The fact that Perplexity's system copies Plaintiffs' content even when not specifically asked to do so further reinforces that Perplexity's system is not "automatic"—it makes an active, unprompted choice to copy, retrieve, and repackage Plaintiffs' content.

### B.    Plaintiffs Are Not Required to Plead All Possible Examples of Perplexity's Infringing Outputs.

Plaintiffs' allegations, which include many specific examples of Perplexity's infringement, surpass the bar for stating a claim for copyright infringement. Perplexity seeks to recast this abundance as a deficit to manufacture a new pleading standard that requires Plaintiffs to allege every single example of every single instance of infringement. *See* Mot. at 12. No rule or case requires such allegations. Nor could it, because "the true extent of" Perplexity's copyright infringement "is a fact uniquely in the possession of Perplexity." Compl. at ¶ 65.

"To survive dismissal, the complaint must allege: (1) which specific original works are the subject of the copyright claim, (2) that plaintiff owns the copyrights in those works, (3) that the copyrights have been registered in accordance with the statute, and (4) by what acts and during what time the defendant infringed the copyright." *Energy Intelligence Group, Inc. v. Jefferies, LLC*, 101 F. Supp. 3d 332, 338 (S.D.N.Y. 2015) (internal citations omitted). Perplexity does not dispute that Plaintiffs have sufficiently identified their copyrighted works and have pled that Plaintiffs own those registered copyrights.

With respect to the fourth element, Plaintiffs are required only to explain *how* Perplexity infringed their copyrighted works. *See Warren v. John Wiley & Sons, Inc.*, 952 F. Supp. 2d 610, 618 (S.D.N.Y. 2013) (plaintiff must provide "sufficient notice to [defendant] as to how the copyrights at issue have been infringed," but is not required to "specify how each *particular*

16

photograph has been infringed" (emphasis added)). Plaintiffs did so in detail. Plaintiffs allege that Perplexity infringes Plaintiffs' copyrighted content when the RAG process copies Plaintiffs' content and uses that content to produce outputs that "contain and/or are derived from Plaintiffs' copyrighted content." Compl. at ¶¶ 114–15. "Perplexity has copied hundreds of thousands of Plaintiffs' copyrighted articles for its RAG database without authorization." *Id*. at ¶ 65. In response to a user prompt, Perplexity produces output using its RAG process, in which the platform "obtain[s] and cop[ies] content from its search index relating to the prompt," "combine[s] the original prompt with the retrieved copied content in order to provide additional context," and "provide[s] the combined data to an LLM, which generates a natural-language response." *Id*. at ¶ 49. The answers that Perplexity generates "sometimes contain full or partial verbatim reproductions of Plaintiffs' copyrighted articles. At other times, Perplexity's answers are reworded into text that resembles, paraphrases, or summarizes Plaintiffs' copyrighted works." *Id*. at ¶ 72. And "[e]ven when the user does not prompt Perplexity to limit its answers to Britannica's content, Perplexity often does so." *Id*. at ¶ 77. "Perplexity's output answers are designed to, and do, act as substitutes for users clicking on links and otherwise going to Plaintiffs' own websites to read Plaintiffs' copyrighted work, thus diverting revenue away from Plaintiffs to Perplexity." *Id*. at ¶ 73. These detailed allegations meet the pleading requirement.

Perplexity wants to re-write the fourth pleading requirement—"by what acts and during what time the defendant infringed the copyright"—to instead require plaintiffs to provide specific examples of every instance of infringement. This court has repeatedly rejected that proposition. For example, in *Energy Intelligence Group, Inc. v. Jefferies, LLC*, 101 F. Supp. 3d 332 (S.D.N.Y. 2015), this court held that the plaintiffs stated a claim by alleging that copyrighted publications subject to user restrictions were "regularly downloaded from Plaintiffs' website" on several

17

different computers affiliated with the defendant, all using the same user login. *Id.* at 340. "To require more at this stage is to require evidentiary detail." *Id.*; *see also Lefkowitz v. McGraw-Hill Glob. Educ. Holdings, LLC*, 23 F. Supp. 3d 344, 354 (S.D.N.Y. 2014) ("For a portion of the [copyrighted works], Plaintiff identifies the publication and information regarding the license limits. The fact that Plaintiff did not include this information for all instances of infringement does not render the [complaint] insufficient, because Plaintiff need not include these allegations in order to plead his claim for copyright infringement adequately." (internal citation omitted)); *Schneider v. Pearson Educ., Inc.*, 2013 WL 1386968, at *3 (S.D.N.Y. Apr. 5, 2013) (rejecting argument that plaintiff was required to plead the specific infringement that occurred with respect to each particular photograph and holding that plaintiff stated a claim by alleging that the photographs were published "without permission, without valid permission being obtained prior to copying, or in excess of its limited licenses"); *Elektra Entm't Group, Inc. v. Barker*, 551 F. Supp. 2d 234, 238–39 (S.D.N.Y. 2008) (holding that plaintiff stated a claim by alleging that the defendant "has continued to use, an online media distribution system to download, distribute, and/or make available, copyrighted works belonging to Plaintiffs").

The fact that Plaintiffs *exceeded* the pleading requirements by including certain specific examples of outputs does not heighten the pleading standard, especially where Plaintiffs specifically alleged that the examples given "are merely a few examples of Perplexity generating verbatim or near-verbatim text and identical curation of content" from Britannica. Compl. at ¶ 78. There is no requirement that the Complaint recite every single potentially infringing output, and Perplexity cites no authority to support this position.

Perplexity then argues that the lack of specific examples as to all fourteen copyrights that Plaintiffs appended to the Complaint means that the Court cannot assess the "substantial

18

similarity" between Plaintiffs' copyrighted content and Perplexity's unauthorized copies. *See* Mot. at 13–14. This argument both ignores Plaintiffs' allegations and misconstrues the caselaw. First, in four different instances, the Complaint lays out Plaintiffs' content and Perplexity's unauthorized reproductions side by side. *See* Compl. at pp. 30 (definition of "plagiarize"), 32–35 (Druids), 37 – 39 (quantum physics), and 41–42 (mysterious disappearances). The substantial similarity between Plaintiffs' content and Perplexity's copy is undeniable in these examples.

Second, Perplexity misunderstands that "substantial similarity" is often a question of fact reserved until after discovery, and can be determined on the pleadings only in particular circumstances not present here. In *Peter F. Gaito Architecture, LLC v. Simone Dev. Corp.*, 602 F.3d 57 (2d Cir. 2010), the Second Circuit distinguished between cases in which the question of substantial similarity is "reserved for resolution by a jury" and others where a district court can resolve that question "as a matter of law." *Id.* at 63. Specifically, where "the works in question are attached to a plaintiff's complaint, it is entirely appropriate for the district court to consider the similarity between those works in connection with a motion to dismiss, because the court has before it all that is necessary in order to make such an evaluation." *Id.* at 64. By contrast, in other circumstances, "the question of substantial similarity cannot be addressed without the aid of discovery or expert testimony." *Id.* at 65. Notably, the Second Circuit cited a copyright infringement case involving "computer programs" as one in which substantial similarity cannot be addressed without fact and expert discovery. *See id.* (citing *Computer Assocs. Int'l, Inc. v. Altai, Inc.*, 982 F.2d 693, 713 (2d Cir. 1992)).

Here, Plaintiffs allege that "Perplexity has copied hundreds of thousands of Plaintiffs' copyrighted articles for its RAG database without authorization" and that "the true extent" of the copying "is a fact uniquely in the possession of Perplexity." Compl. at ¶ 65. Plaintiffs similarly

19

allege that "since it was launched in 2022, Perplexity included all of Plaintiffs' copyrighted content

as RAG Content," "including by scraping content from Plaintiffs' websites." *Id.* at ¶ 52. Plaintiffs

could not reasonably have attached these hundreds of thousands of works in question to the

Complaint. Further, to the extent that an understanding of Perplexity's RAG process is relevant to

assess substantial similarity between Plaintiffs' content and Perplexity's output, expert opinion

may be "warranted" to explain "the highly technical nature of computer programs." *See Computer*

*Assocs. Int'l*, 982 F.2d at 713.

Perplexity inaptly compares this case to infringement disputes regarding discrete works

that can be reproduced in full with the complaint. It relies exclusively on cases involving creative

works such as television plots and pop songs for which the district court can make a substantial

similarity finding at the 12(b)(6) stage because the complaint and attachments, standing alone,

contain "all that is necessary" to compare the works. *See Piuggi v. Good for You Productions LLC*,

739 F. Supp. 3d 143, 152 (S.D.N.Y. 2024) (cited in Mot. at 12) (two television shows' alleged

copying of reality television plot and concept); *Shull v. TBTF Productions Inc.*, 2019 WL 5287923,

at *1 (S.D.N.Y. Oct. 4, 2019) (cited in Mot. at 14) (one television show's alleged copying of

premise and characters from one book); *Abdin v. CBS Broad. Inc.*, 971 F.3d 57, 60 (2d Cir. 2020)

(cited in Mot. at 14) (one television show's alleged copying of premise and character from one

video game); *Hines v. Roc-A-Fella Records, LLC*, 2020 WL 1888832, at *1 (S.D.N.Y. Apr. 16,

2020) (cited in Mot. at 13) (two modern songs' alleged copying of one older song); *Larball Publ'g*

*Co., Inc. v. Lipa*, 2023 WL 5050951, at *1 (S.D.N.Y. Aug. 8, 2023) (cited in Mot. at 13) (one

modern pop song's alleged copying of two older songs). Perplexity also cites three cases involving

purported artwork or screenplays in which the plaintiffs did not provide *any* specific allegations

regarding how or in what way the defendants' movies purportedly copied the works. *See Adams*

*v. Warner Bros. Pictures Network,* 2005 WL 3113425, at \*1 (E.D.N.Y. Nov. 22, 2005) (cited in

Mot. at 12–13); *Evans v. NBCUniversal Media, LLC*, 2021 WL 4513624, at \*5 (C.D. Cal. July 23,

2021) (cited in Mot. at 13); *Stevens v. Tomlin*, 2023 WL 5486247, at \*1 (E.D.N.Y. Aug. 24, 2023)

(cited in Mot. at 14). These cases provide no ground to dismiss Count II of Plaintiffs' Complaint.

## V.     CONCLUSION

Plaintiffs respectfully request that the Court deny Perplexity's Motion to Dismiss Count II

of the Complaint. Dkt. 30. In the alternative, Plaintiffs request that the Court grant Plaintiffs leave

to amend the Complaint. *Loreley Fin. (Jersey) No. 3 Ltd. v. Wells Fargo Sec., LLC*, 797 F.3d 160,

191 (2d Cir. 2015).

Dated: November 24, 2025                    */s/  Ian B. Crosby*

Ian B. Crosby (*pro hac vice*)
SUSMAN GODFREY L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
icrosby@susmangodfrey.com
T: (206) 516-3880
F: (206) 516-3883

Davida Brook *(pro hac vice)*
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400
Los Angeles, CA 90067
dbrook@susmangodfrey.com
T: (310) 789-3100
F: (310) 789-3150

Y. Gloria Park
Sarah Hannigan
SUSMAN GODFREY L.L.P
One Manhattan West, 50th Street
New York, NY 10001
gpark@susmangodfrey.com
shannigan@susmangodfrey.com
T: (212) 336-8330

21

F: (212) 336-8340

*Attorneys for Plaintiffs*
*Encyclopaedia Britannica, Inc.*
*and Merriam-Webster, Inc.*

## **CERTIFICATE OF SERVICE**

I certify that on this 24th day of November, 2025, I caused a true and correct copy of the foregoing to be electronically filed with the Clerk of the Court using the CM/ECF system which will send notification to the attorneys of record.

/s/ *Ian B. Crosby*
Ian B. Crosby

## **CERTIFICATE OF COMPLIANCE WITH WORD COUNT**

I certify that Plaintiffs' Memorandum of Law in Opposition to Perplexity's Motion to Dismiss satisfies the word count requirement in Local Civil Rule 7.1(c). Plaintiffs' Memorandum is 6,619 words, excluding the exempt categories described in Rule 7.1(c).

/s/ *Ian B. Crosby*
Ian B. Crosby

23