EXHIBIT A

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates to: *All Cases*	Case No.: 1:25-md-3143 (SHS) (OTW) Hon. Sidney H. Stein

[PROPOSED] BRIEF OF AMICUS CURIAE JIM ALBRECHT IN SUPPORT OF NEWS PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT

TABLE OF CONTENTS

I. INTRODUCTION 1

II. INTEREST OF AMICUS CURIAE 1

III. ARGUMENT 2

A. BILLIONS OF COPIES. 2

B. IMPLIED LICENSES ARE NOT AN UNSCOPED PERMISSION. 3

C. A LICENSE SHOULD NOT BE IMPLIED FOR USES THAT BLOCK THE CONTENT OWNER'S EXPECTED BENEFIT. 4

D. IMPERMISSIBLE COPYING OF NEWS FACTS. 6

E. OPENAI IS BEING ATTRACTIVELY COY: THEY MADE A GREAT NEWS PRODUCT. 7

F. THE ECONOMIC HARM IS ONGOING AND INEVITABLE. 9

G. THE COPYING IS THE VIOLATION; THE OUTPUT DEMONSTRATES THE SUBSTITUTIVE PURPOSE. 11

H. THE COPYING IS SUBSTANTIAL. 12

IV. CONCLUSION 13

TABLE OF AUTHORITIES

Page(s)

Cases

Assessment Techs. of WI, LLC v. WIREdata,
350 F.3d 640 (2003) 6

Associated Press v. Meltwater U.S. Holdings, Inc.,
931 F. Supp. 2d 537 (S.D.N.Y. 2013) 6, 7

Bitmanagement Software GmBH v. United States,
989 F.3d 938 6

Est. of Hevia v. Portrio Corp.,
602 F.3d 34 (1st Cir. 2010) 3

Feist Publications, Inc. v. Rural Tel. Serv. Co.,
499 U.S. 340 (1991) 11

Field v. Google Inc.,
412 F. Supp. 2d 1106 (D. Nev. 2006) 3

Harper & Row Publishers, Inc. v. Nation Enters.,
471 U.S. 539 (1985) 6

John G. Danielson, Inc. v. Winchester-Conant Props., Inc.,
322 F.3d 26 (1st Cir. 2003) 3

Miller v. Universal City Studios, Inc.,
650 F.2d 1365 (5th Cir. 1981) 6

Muhammad-Ali v. Final Call, Inc.,
832 F.3d 755 (7th Cir. 2016) 3, 5

Nelson-Salabes, Inc. v. Morningside Dev., LLC,
284 F.3d 505 (4th Cir. 2002) 3

Sega Enters. Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992) 6

Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc.,
No. 25-2153, 2026 WL 2932557 (3d Cir. Sept. 29, 2026) 12

I. INTEREST OF AMICUS CURIAE

Amicus Jim Albrecht served as Senior Director of News Ecosystem Products at Google from 2017 to 2023. In that role, he developed products intended to strengthen the online environment for news publishers and enhance the value exchange between publishers and search platforms. He has also publicly addressed, including in The Washington Post, the harms that generative search may cause to the news ecosystem.[1]

Amicus has a substantial interest in this case because it concerns the use of news content in retrieval-augmented generation systems and the economic consequences of that use for publishers. Drawing on his experience with search products, web-traffic patterns, and the relationship between online platforms and news organizations, Amicus seeks to assist the Court in understanding how these systems retrieve and copy publishers' content, how they differ from traditional search engines, and how generated answers may substitute for visits to publishers' websites and thereby diminish traffic, advertising revenue, subscription opportunities, and other commercial benefits.

II. INTRODUCTION

By almost any measure, I have spent way too much time thinking about the "news ecosystem"—i.e., that jumble of websites, search platforms, public data, social websites, blogs, and influencers in which consumers find information and content owners find customers—and even had that term in my job title at Google for many years. My work in that role focused on building products that made the web a more fruitful environment for news publishers, and that enhanced the value exchange that took place in the symbiotic relationship between publishers and Google search in particular. But in the years since ChatGPT introduced the world to LLM-based

[1] Jim Albrecht, The real wolf menacing the news business is AI, Wash. Post (Feb. 6, 2024), https://www.washingtonpost.com/opinions/2024/02/06/ai-news-business-links-google-chatgpt/

chatbots, with their voracious appetite for both user attention and information with which to keep that attention rapt, the ecosystem has changed dramatically. For one thing, that shift in the flow of user attention has been followed (as ever) by a shift in the flow of economic value. For another, both of those changes have ridden a flood tide of "intermediate copying"—an industrial scale content reproduction operation called "retrieval augmented generation," or RAG.

III. ARGUMENT

A. BILLIONS OF COPIES.

Retrieval augmented generation is a fancy way of saying "I gave my LLM a search engine." That is, a RAG system lets an LLM search for information related to a user's query and use it to write an accurate answer.[2] The RAG search functions much like the search engines that we are all accustomed to, with one big difference. When you and I use search engines, the software finds documents that contain the information we need and then gives us back search results: a title, 20-30 words of text from the documents, and a link to the web URL for each; when an LLM uses a RAG system, the system copies the relevant documents or a large chunk from each (a couple hundred words, say) and hands them back to the LLM. The end user never sees those copies because the LLM rewrites them in the answer it provides to the user. But the copies are made nevertheless and are distributed to the LLMs for each of 100s of millions of ChatGPT sessions every day.[3]

Of course, this sort of copying—computers making copies inside computer infrastructure in order to facilitate things the computers are doing—has gone on long before there were LLMs and RAG systems. Web browsers themselves make local copies of texts on remote servers so that

[2] It is important to note that, as a search facility used by an LLM, RAG is completely unrelated to LLM training, uses content in a completely different way, and raises a distinct set of copyright and business practice questions.

[3] ChatGPT reportedly serves billions of chat sessions per day and public estimates suggest one third of these are supported by RAG operations.

users can read them on their personal devices. And search engines, for instance, have always copied documents verbatim in order to figure out which ones were relevant to a user's query and then present the user with links to those documents.

B. IMPLIED LICENSES ARE NOT AN UNSCOPED PERMISSION.

Most of this computer copying has been ignored in the past or considered to be covered by implied licenses, either because it was ephemeral and without economic consequence (e.g., people do not even reflect on the invisible fact that copies are made whenever documents are loaded from a hard drive into memory) or because the copying was instrumental to the content owners' own economic interest in harvesting the value of their own content.

An implied nonexclusive copyright license generally exists when the licensee requests creation of a work, the licensor creates and delivers that work to the requesting licensee, and the licensor intends that the licensee copy and distribute or otherwise use it. *See Muhammad-Ali v. Final Call, Inc*., 832 F.3d 755, 762 (7th Cir. 2016); *see also Est. of Hevia v. Portrio Corp*., 602 F.3d 34, 41 (1st Cir. 2010) (explaining that the three-part implied-license inquiry ultimately turns on the copyright owner's intent to permit the work's use). It is the final criterion: the intention of the licensor that the licensee copy and distribute the work, which is the determinative consideration. *See Muhammad-Ali*, 832 F.3d at 762; *see also John G. Danielson, Inc. v. Winchester-Conant Props., Inc.*, 322 F.3d 26, 40 (1st Cir. 2003) ("the touchstone for finding an implied license, according to this framework, is intent.") (*citing Nelson-Salabes, Inc. v. Morningside Dev., LLC*, 284 F.3d 505, 515 (4th Cir. 2002). Often, a license may be implied from conduct from which a user could properly infer intent, and silence can manifest consent where the copyright owner knows of and encourages the use. *See Field v. Google Inc*., 412 F. Supp. 2d 1106, 1116 (D. Nev. 2006) (holding that Google did not directly infringe the plaintiff's copyrighted works by operating its cache and displaying "cached" links). For example, publishers' websites

are worthless to them if the browser does not present the sites' content to their subscribers; and the search engine cannot promote publisher website links to potential new customers without first copying and indexing the content.

But to say that such implied licenses cover this style of copying for any and all purposes, irrespective of how far they stray from the intended use case, would require an astonishing unwillingness to make distinctions. Such is especially true of RAG's prolific copying, the circumstances of which are so clearly distinct from these other types of intermediate copying.

First and most obviously, having a computer visit your website and copy your documents is very different from having a human person do so for the simple reason that advertisers pay for the human attention generated by the content, not for computer attention. And if the purpose of the computer's valueless visit is not to create a search product that promotes your brand and content, but to copy your document and distribute it to another computer that in turn recapitulates its content for a user who was looking for exactly that sort of information…well, then the copying comes not merely without benefit, but with significant cost.

C. A LICENSE SHOULD NOT BE IMPLIED FOR USES THAT BLOCK THE CONTENT OWNER'S EXPECTED BENEFIT.

It's understandable that companies developing artificial intelligence find it irresistible—almost compulsive—to talk about how much like human users their products are. But they become strangely mute when it comes to the negative implications of such comparisons. Here's how OpenAI says it handles a user query:[4]

> ChatGPT would send that query to Microsoft Bing, which returned a list of results with links to relevant webpages, along with short "snippets" of text, similar to what a user would see if they ran the search on www.Bing.com. ChatGPT analyzed the search results returned by Bing to determine which webpages might contain

[4] OpenAI neglects to mention that it stores a copy of the document it retrieves, such that it can make additional local copies of it for every subsequent session without visiting the website at all.

> relevant information, just as a user may scan the search results, and then ChatGPT sent an automated request to websites for the content of those webpages. After receiving the content, ChatGPT would extract the relevant factual information and synthesize it into a response to the user.[5]

It's an interesting set of reflections: OpenAI's LLM takes the place of the human user in the use of a search engine, in the evaluation of its results, in the visit to the content owner's website, and in the extraction of the information of interest—that is to say, all the parts where having an actual human present generates economic value for the content owner. That lost value starts with brand discovery in search; continues with visits to the publisher's website, ad impressions, recirculation to other features, and subscriptions; and ends with the deeper brand connection that comes from reading and appreciating the specific value of the content on the specific brand.

OpenAI's ability to capture each of these parts of the publisher's economic value funnel relies on copying its content: the copying the search engine does to generate its results; the copying of the document for analysis by the LLM; and the copying of specific chunks of the document for re-writing by the LLM as it synthesizes a response. Yes, OpenAI is right: the LLM performs many human activities in this process, and by doing so it saves the actual human quite a bit of time—so much that users will pay OpenAI for its services. And it would be silly to say that it ought not be allowed to build these labor-saving products. Equally fatuous, however, is the contention that it should continue to benefit from implied content licenses underpinning web browsers and search engines having built a product that takes for itself all the economic value those licenses were predicated on.

Intent determines the existence of implied licenses. The primary focus is on the licensor's objective intent at the time of creation and delivery. *See Muhammad-Ali*, 832 F.3d at 762; *see also*

[5] Def.'s Mem. P. & A. Supp. Mot. Summ. Judg. ECF No. 1496, pg. 13.

Bitmanagement Software GmbH v. United States, 989 F.3d 938, 953–55 (Newman, J., concurring) (Fed. Cir. 2021) (the court considered what economic purpose copyrighted software served and what rational basis plaintiff could have had for consenting to unlimited copying). How can it seriously be said that content owners objectively intend for LLMs to copy large chunks of their websites to no economic advantage of the content owner, indeed, likely at an economic disadvantage to the content owner?

D. IMPERMISSIBLE COPYING OF NEWS FACTS.

For better or for worse, the bare facts of the news receive limited copyright protection. *See Harper & Row Publishers, Inc. v. Nation Enters*., 471 U.S. 539, 556 (1985) (no author may copyright the factual information he narrates). However, the form, expression, arrangement, and selection of these facts are generally copyrightable. *See Miller v. Universal City Studios, Inc.*, 650 F.2d 1365, 1368 (5th Cir. 1981). This has not changed in the age of the internet. In *Associated Press v. Meltwater U.S. Holdings, Inc.*, 931 F. Supp. 2d 537 (S.D.N.Y. 2013), the Southern District of New York granted summary judgment against a defendant that used automated programs to scrape news articles from the web and redistribute excerpts, ranging from 4.5% to 61% of individual articles.

OpenAI justifies its RAG copying based on *Assessment Techs. of WI, LLC v. WIREdata* 350 F.3d 640, 645 (2003), which suggests that one can claim "fair use" and make an intermediate copy of a copyrighted work if one has no other way to obtain the *factual information* it contains. The disassembly of copyrighted works and extraction of factual information is fair use when it is the only means of accessing elements not protected by copyright and the copier has a legitimate reason for seeking access. *See Sega Enters. Ltd. v. Accolade, Inc*., 977 F.2d 1510, 1518 (9th Cir. 1992), as amended (Jan. 6, 1993) (plaintiff's video game programs contained unprotected logic,

structural, and visual display elements that could not be examined without copying, and thus are afforded a lower degree of protection than more traditional literary works).

However, the line between permissible copying of facts and impermissible systematic scraping of news articles is fine. Indeed, to operate, LLMs make practically endless copies of news articles, a process the *Associated Press* case found impermissible. *Associated Press*, 931 F. Supp. 2d.

The entire news industry exists in a permanent, furious tumult of reporting and re-reporting each other's stories. Anyone can enter the fray and do that reporting and re-reporting. There is no prison tower where publishers keep their captive supply of new facts. You can go out and pluck them in the wild just as they do. But you do have to do the work. And the work costs money. In other words, there is no necessity driving OpenAI to take web content and make endless copies. That is just the easiest and cheapest way to get it.

E. OPENAI IS BEING ATTRACTIVELY COY: THEY MADE A GREAT NEWS PRODUCT.

OpenAI rejects the idea that its grounded chatbot could substitute for a news product, writing: "Browse [OpenAI's RAG facility] finds and analyzes publicly available online content, extracts facts, and generates an original response with citations to relevant resources." Understanding exactly how this is supposed to distinguish Browse from a reporter's work is less important than understanding that it fails to do so. This is less a defect in expressing the point than in its fundamental credibility. It's very challenging indeed to build the logic for the case that a product that inexhaustibly answers questions about the day's news events cannot be considered a news product. To be fair, LLM and search platforms have long struggled to acknowledge that they inherently compete with information verticals. This becomes increasingly true as a card-catalog-style search engine displaying links to documents gradually transforms into an answer engine.

Search for a stock symbol on Bing, for instance, and you will not get links to Yahoo Finance and Charles Schwab at the top of the page; you get Microsoft's own Bing Finance product—which is quite a nice product but which obviously competes with other finance products.

These first vertical features in search engines tended to be in areas that already had commodified data suppliers (e.g., sports and finance), which meant that the data could be acquired easily and relatively cheaply. It also meant that the search product could enter the vertical space without undue legal risk. Imagine, on the contrary, if Bing Finance were obtaining stock data by crawling or scraping: it would violate the terms of service of the intermediate licensees and also risk charges of misappropriation by the licensor.

But when you take a search product and put a chatbot in front of it, which then uses it to research and narrate all answers in any arena, you are now squarely in the business of…being in every vertical information business. You may have just done it without the caution and legal scruples of prior search products.

The nicest thing I can say on this point is that it's admirably modest for OpenAI to claim that ChatGPT cannot substitute for a news article or compete with a news site, but the argument seems to rely on one never having used their product. After all, ChatGPT can provide a comprehensive, tailored response to a user's specific expressed interest and then can be interrogated further for deeper or adjacent information in the same article and others. Users can continue the interrogation (and indeed the chatbot encourages it) until they have exhausted all the information in those articles and moved on, locust-like, to row after row of the sweet corn of human knowledge.

For a certain segment of the news audience, it's not just a competitive news product; it's the perfect news product, the one they've always wanted: one that lets you skip reading all the

articles and have your assistant narrate only the parts you're curious about. OpenAI has managed to create an idealized form of the personalized website, and they are right to be proud of their accomplishment. But that astonishing product works, in no small part, because it has access to the endless copies its RAG facility makes of news articles and documents from all other verticals.

F. THE ECONOMIC HARM IS ONGOING AND INEVITABLE.

In short, no reasonable person can use ChatGPT or Microsoft Copilot and fail to see its capacity to replace news products, *at least for some significant part of its audience*. This principle of fractional loss of commercial value is a well-understood principle of web traffic flows dating all the way back to the beginning of the commercial web in the mid 1990s. As a young man, I managed a celebrity site called *Mr Showbiz*, which presented users with celebrity news and other features suitable (cough) to fans of pop culture. We spent a lot of time analyzing data to understand why people clicked on some things and not others. Beyond the basic allure, attractiveness, or sensationalism of the topics, one obvious factor was the amount of content provided before the click. Beyond providing a topically clear and engaging headline, almost every added piece of information diminished click-through.

In the 2000s, this basic insight engendered a whole sub-industry of "clickbait" websites. These sites specialized in finding the one detail of the story that would spark the most curiosity in the mind of the user, and make the click the price of satisfying that need to know. (e.g., "What two burglars found in this home scared them so badly they turned *themselves* in!") At some level, all attention-based media business is like this: there is the question and the answer; the joke and the punchline. Reducing the anticipation in any way reduces the click-through.

This is equally true in search. Studies have found that search result types with longer snippets reduce total referral traffic.[6] More recently, AI Overviews in Google search have reportedly reduced clicks by 58% according to a study by ahrefs.com.[7] And the News Plaintiffs' Combined Summary Judgment Brief has already noted the vastly smaller (83-93% less) referral traffic in Microsoft Copilot vs Bing Search, and has recorded various other relevant data points, which I will not repeat here. I'll merely say that all of this is well-understood conventional wisdom.

So, is it possible that ChatGPT's generated narration of the facts in a news article does not similarly sap the commercial value of the article that it has copied from and then rewritten? No, it most certainly is not. This is a fundamental principle of human behavior, not an attribute of the Microsoft and Google product suites or of sites that use a certain link color. Put information before the click in small, medium, and large amounts, and you will lose audience to small, medium, and large degrees.

OpenAI seems to argue that in order to be seen as competing with news websites, they must reproduce every feature of those products and satisfy every news user's needs, but that is not how any content business works in practice. Rather, they work by aggregating an audience with diverse interests, needs, and levels of engagement. Using copies of the news publishers' content in a way that satisfies the needs of 20 or 40 or 60 or 93% of that audience is not something you can reasonably dismiss as not substitutive merely because it's not yet 100%.

Nor can one escape the reality of having launched a competitive and substitutive product merely by gesturing at more grandiose ambitions. Yes, it may seem a small thing to complain that

[6] Tim Soulo, Ahrefs' Study Of 2 Million Featured Snippets: 10 Important Takeaways, Ahrefs Blog (May 29, 2017), https://ahrefs.com/blog/featured-snippets-study/.

[7] yan Law, Update: AI Overviews Reduce Clicks by 58%, Ahrefs: Blog (Feb. 4, 2026), https://ahrefs.com/blog/ai-overviews-reduce-clicks-update/.

OpenAI's RAG system and its massive real-time copying operation is continuously siphoning value from a class of underlying products (there's a reason why they call it "grounding") to produce one that substitutes for all of them, when the real goal is to substitute for human intelligence itself. Nevertheless, intellectual property rights must be respected even when they inconvenience men (and products) of destiny.

G. THE COPYING IS THE VIOLATION; THE OUTPUT DEMONSTRATES THE SUBSTITUTIVE PURPOSE.

OpenAI will hasten to say that its product cannot violate publisher copyright because the texts it generates are no different from a *human's* permissible rewriting of an original work or a *human's* synthesizing of several texts into one. *See Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 344 (1991) (subsequent compiler of facts can use the facts themselves so long as the new work does not replicate the same original selection and arrangement of facts). That is to say, the computers are doing something that courts have found to be transformative and thus a fair use when humans do it; so why should their poor computer be any different?

Let us stipulate their point and inspect what appears to be an under-examined implication of the analogy: namely, they are taking a copyrighted text, reproducing it without authorization, and passing those copies to dozens or hundreds or thousands or *millions* of these notional employees—practically just like us humans, they say—who are tasked with generating commercial rewrites. Would the court look sympathetically on a company making and distributing thousands of unauthorized copies of a protected work to its employees so that they might rapidly recapitulate that work for commercial purposes and by so doing reduce demand for the original?

The analogy forces us to look squarely at the fact that the RAG system plays a trick on fair use standards that no one could have foreseen 100 years ago: it reduces the prohibitive cost and latency of wide-scale human rewriting of content to fractions of pennies and fractions of seconds

by training the computer to act like a human. And the consequence of these pseudo-human workers is that people who research, assemble, verify, and publish new facts are severed from any reasonable expectation of harvesting the fruits of their own labor. But the trick has a second edge to it, because the training alone does not accomplish the goal: each of the millions of pseudo-human workers—each user session with an LLM—requires a content retrieval capability, its own supply of freshly copied documents. OpenAI wants to have it both ways: the computer re-writing is fair use, just as if a human did it; but the computer's unauthorized copying and distribution to these virtual humans is an internal computer operation we can ignore, even though it, too, is exactly what a human would do: you would make copies of the thing you want re-written and pass them out to your re-writers. But you would need a license for that.

In this respect, the case is very much like *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc.,* No. 25-2153, 2026 WL 2932557 (3d Cir. Sept. 29, 2026). In both cases, a company built a product that satisfies the same type of information needs as an existing product and extensively copied the existing company's content—not to republish it, but to use it in its own system to achieve a quality product. In the ROSS case, that meant successfully answering questions about the law; here, it means successfully answering questions about the news. One difference is that ROSS used the copies in a training step that was completed at a point in time. RAG usage continues generating new copies daily and indefinitely.

H. THE COPYING IS SUBSTANTIAL.

RAG systems retrieve variable amounts of verbatim text, but the practice generally requires a couple hundred words to get enough context to have meaningful confidence in the accuracy of what gets generated. With standard news articles landing at ~800-1000 words nowadays, that means a very significant portion of the source document will be copied verbatim and distributed

to the computer worker for rewriting in each pertinent session. So already, the copying is substantial.

From the rights holder's perspective, the reality is worse; over time, the document will be retrieved for use in an indeterminately large number of user questions and follow-up questions. Thus, across that many sessions, as a computer worker receives a user query and retrieves a particular hundreds-of-words chunk, the entire document may end up being reproduced many, many times. The RAG system is a wholesale verbatim copying mechanism that is adept at parceling out to the workers the particular chunks they need to rewrite in order to achieve the company's commercial goals.

In other words, it serves as no real defense in the context of the above analogy if I were to distribute my thousands of unauthorized copies to my thousands of rewriters in *only* 10-40% chunks. I still made the unauthorized copies of the entire text, and I still did it with commercial intent. I merely spread out the copying across sessions and workers. Suffice it to say that in the business of sneaking off with someone else's property, it's no saving virtue to take your time about it. Though I can certainly see how it would save compute costs.

IV. CONCLUSION

Law is supposed to help us find a reasonable path to justice amid the thickets of competing claims and the tangle of obfuscating or even disingenuous applications of traditional standards & heuristics. Reasonable people might see the status of RAG systems in two very different lights: one might see the copying and generative work these computing systems do as novel and demanding reassessment of the traditional tests of fair use and of their viability on the other side of this technological revolution; another might treat the operation of these systems as simply and properly analogous to the activity of human workers under a still-robust fair-use standard. But what no reasonable person can do is treat as human the computers' generated transformations of

these texts for a competitive product, while ignoring as ephemeral and phantom the incessant, unauthorized copying on which that work depends.

Dated: New York, NY
October 30, 2026

Respectfully submitted,

By: _____

Christopher M.W. Pioch, Esq.
Christopher Pioch and Associates
1350 Avenue of the Americas, Fl 2-1075
New York, NY 10019
Tel: (201) 956-0040
E: cpioch@piochlaw.com

Attorney for Proposed Amicus Curiae Jim Albrecht

CERTIFICATE OF WORD COUNT COMPLIANCE

In accordance with Local Civil Rule 7.1(c) and the Court's August 7, 2026 Order (Dkt.1663, 25-md-3143), I certify that the foregoing amicus brief contains 4,373 words, including footnotes but exclusive of the caption page, table of contents, table of authorities, certificates, and signature block. The basis of my knowledge is the word count feature of the word-processing system used to prepare this brief.

Dated: New York, NY
October 30, 2026



Christopher M.W. Pioch