UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

In Re: OpenAI, Inc. Copyright Infringement Litigation This Document Relates To: Case No. 1:23-cv-11195 Case No. 1:24-cv-03285 Case No. 1:24-cv-04872	Case No. 1:25-md-3143-SHS-OTW **REDACTED – PUBLIC VERSION**

Defendant Microsoft Corporation's Memorandum of Law In Support Of Motion For Summary Judgment In News Plaintiffs' Consolidated Cases

TABLE OF CONTENTS

Page

TABLE OF AUTHORITIES iii
TABLE OF CITATION CONVENTIONS vi
ORIENTATION 1
INTRODUCTION 1
STATEMENT OF FACTS 4
Copilot pairs LLM technology with search infrastructure to create a powerful new knowledge and information tool. 4
Web grounding uses retrieval augmented generation, or RAG, to ground responses in external webpages drawn from the Bing search index. 5
Microsoft designed Copilot in accord with its commitment to responsible AI. 7
News outlets claim that training LLMs and web grounding are copyright infringement. 9
ARGUMENT 10
I. Training An LLM Is Fair Use As A Matter Of Law. 10
II. Copilot Web Grounding Is Fair Use As A Matter Of Law. 13
A. Copilot web grounding creates a new product for finding and using information (Factor one). 14
B. Intermediate grounding copies use factual elements and do not display protected expression to users (Factors two and three). 18
1. News articles are predominantly factual. 18
2. Grounding copies are intermediate versions made for purposes of influencing generated responses. 19
3. Copilot responses displaying incidental verbatim fragments of News Plaintiffs' works are exceptionally rare. 20
C. Web grounding is not a substitute for News Plaintiffs' works in any cognizable market (Factor four). 23
1. Plaintiffs cannot demonstrate harm in any traditional market. 23
2. The Court should not recognize a market for grounding content. 28
3. News Plaintiffs are not entitled to entrench their business models. 29
III. News Plaintiffs' Knowing Decision To Forego Use Of Copilot Meta Tags Gives Rise To An Implied License. 31

IV. At Most, News Plaintiffs Are Entitled To One Award Of Statutory Damages Per Issue. 34
CONCLUSION 35
CERTIFICATE OF WORD COUNT COMPLIANCE 39

TABLE OF AUTHORITIES

Page(s)

Cases

ABKCO Music, Inc. v. Sagan,
50 F.4th 309 (2d Cir. 2022) 31

Am. Geophysical Union v. Texaco Inc.,
60 F.3d 913 (2d Cir. 1994) 29

Associated Press v. Meltwater U.S. Holdings, Inc.,
931 F. Supp. 2d 537 (S.D.N.Y. 2013) 33, 34

Authors Guild, Inc. v. HathiTrust,
755 F.3d 87 (2d Cir. 2014) 10, 11

Authors Guild v. Google, Inc.,
804 F.3d 202 (2d Cir. 2015) 11, 16, 18, 19, 20, 21, 22, 23, 28

Bartz v. Anthropic PBC,
787 F. Supp. 3d 1007 (N.D. Cal. 2025) 10

Bryant v. Media Right Prods., Inc.,
603 F.3d 135 (2d Cir. 2010) 35

Capitol Recs., LLC v. ReDigi Inc.,
910 F.3d 649 (2d Cir. 2018) 12

Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc.,
150 F.3d 132 (2d Cir. 1998) 29

Consumers Union of U.S., Inc. v. Gen. Signal Corp.,
724 F.2d 1044 (2d Cir. 1983) 18

EMI Christian Music Grp., Inc. v. MP3tunes, LLC,
844 F.3d 79 (2d Cir. 2016) 35

Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.,
499 U.S. 340 (1991) 18

Field v. Google Inc.,
412 F. Supp. 2d 1106 (D. Nev. 2006) 32

Fox News Network, LLC v. TVEyes, Inc.,
883 F.3d 169 (2d Cir. 2018) 17

Google LLC v. Oracle Am., Inc.,
593 U.S. 1 (2021) 13, 14, 16, 17, 28, 30

Hanagami v. Epic Games, Inc.,
85 F.4th 931 (9th Cir. 2023) 21

Harper & Row Publishers, Inc. v. Nation Enters.,
471 U.S. 539 (1985) 23

Infinity Broad. Corp. v. Kirkwood,
150 F.3d 104 (2d Cir. 1998) 11, 23, 28

Kadrey v. Meta Platforms, Inc.,
788 F. Supp. 3d 1026 (N.D. Cal. 2025) 10

Kelly v. Arriba Soft Corp.,
336 F.3d 811 (9th Cir. 2003) 6

N.Y. Mercantile Exch. v. IntercontinentalExchange, Inc.,
497 F.3d 109 (2d Cir. 2007) 13

N.Y. Times Co. v. Microsoft Corp.,
777 F. Supp. 3d 283 (S.D.N.Y. 2025) 9

Nihon Keizai Shimbun, Inc. v. Comline Bus. Data, Inc.,
166 F.3d 65 (2d Cir. 1999) 18

On Davis v. Gap, Inc.,
246 F.3d 152 (2d Cir. 2001) 11

Parker v. Yahoo! Inc.,
No. 07-2757, 2008 WL 4410095 (E.D. Pa. Sept. 25, 2008) 32, 33

Ringgold v. Black Ent. Television, Inc.,
126 F.3d 70 (2d Cir. 1997) 22

Romanova v. Amilus Inc.,
138 F.4th 104 (2d Cir. 2025) 18

Sega Enters. Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992) 20

Sony Comput. Ent., Inc. v. Connectix Corp.,
203 F.3d 596 (9th Cir. 2000) 20

Swatch Grp. Mgmt. Servs Ltd. v. Bloomberg L.P.,
756 F.3d 73 (2d Cir. 2014) 17, 18

A.V. ex rel. Vanderhye v. iParadigms, LLC,
562 F.3d 630 (4th Cir. 2009) 21

Yamashita v. Scholastic Inc.,
936 F.3d 98 (2d Cir. 2019) 31

Statutes

17 U.S.C. § 101.....34

17 U.S.C. § 107(1).....10, 18

17 U.S.C. § 107(2) 18

17 U.S.C. § 107(3).....18

17 U.S.C. § 107(4) 23

17 U.S.C. § 504(c)(1).....34, 35

Other Authorities

3 *Nimmer on Copyright* § 10.03[A].....31

4 *Nimmer on Copyright* § 13F.14[B] 6

TABLE OF CITATION CONVENTIONS

This brief uses the following citation conventions for record materials filed in support of Microsoft's Motion for Summary Judgment.

Citation	Document
Abib ¶	Declaration of Elbio Abib in Support of Microsoft's News Motion for Summary Judgment
Bird ¶	Declaration of Sarah Bird in Support of Microsoft's News Motion for Summary Judgment
Canel ¶	Declaration of Fabrice Canel in Support of Microsoft's News Motion for Summary Judgment
Gaggara ¶	Declaration of Padma Priya Gaggara in Support of Microsoft's News Motion for Summary Judgment
Hurst Ex.	Declaration of Annette L. Hurst in Support of Microsoft's Motions for Summary Judgment
Lafferty ¶	Declaration of John D. Lafferty, Ph.D. in Support of Microsoft's Motions for Summary Judgment
SUMF ¶	Rule 56.1 Statement of Undisputed Material Facts in Support of Microsoft's News Motion for Summary Judgment
Tucker ¶	Declaration of Catherine Tucker, Ph.D. in Support of Microsoft's News Motion for Summary Judgment

All statutory cites are to Title 17 unless otherwise indicated. All ECF cites are to MDL No. 1:25-md-3143-SHS-OTW unless otherwise indicated.

ORIENTATION

Microsoft recommends that its brief in the Books cases be reviewed first, followed by the News brief below.

INTRODUCTION

Microsoft's Books brief explains why the use of copyrighted works to train an LLM is fair use. That conclusion applies equally to news articles. Whatever the nature of the works or sources from which they are acquired, using them to make an LLM that captures the sweep of human language and knowledge is new and different and far beyond any previously existing use or market. Copyright law does not permit rightsholders to block transformative technologies like LLMs; it encourages such uses on the expectation that rightsholders will adapt and the public will be better off for it. This brief addresses News Plaintiffs' challenge to just one of the ways innovators have harnessed LLM technology and put it in the hands of consumers.

At issue is Microsoft's AI chatbot, Copilot, and in particular a functionality called "web grounding." A chatbot without web grounding is certainly *astounding* for its ability to generate human-like responses to many types of prompts. But a web-grounded chatbot is something early chatbots often were not: *useful*, for a variety of work and personal tasks. Web grounding pairs the chatbot with internet search capabilities (in the case of Copilot, with Microsoft Bing). This enables the LLM that powers Copilot to locate, synthesize, and contextualize information in responding to user queries. For a user, the experience is like having an (artificially) intelligent conversation with the internet itself—asking what it knows, digging deeper with follow-ups, and directing it to organize its knowledge into whatever form is most helpful to the task.

Copilot is not just a groundbreaking product, but also a paradigm for responsible and fair AI. Microsoft firmly believes that LLM-based technology must benefit society broadly, including publishers of digital content. In Copilot, Microsoft carried forward longstanding features of traditional search that ensure the benefits of AI technology are shared. Like a search engine, Copilot outputs provide links directly to the webpages relied upon so users can click through. And like a search engine, Microsoft allows any website owner to exclude their content. This last point bears repeating: Any website owner that objects to web grounding with its content can simply opt out.

News Plaintiffs maintain this lawsuit anyway, seeking to ban a product they can readily avoid. When they filed this case, they alleged—and claimed in their own newspapers—that Copilot would *reproduce* large portions of their articles to users. This dire prediction did not materialize. Over 8.2 million chat logs produced in discovery prove that Copilot virtually never displays even a sentence of News Plaintiffs' content to users. This is unsurprising since web grounding is designed to use back-end, search engine portions of webpages to *influence* the Copilot LLM's *generated* response to a user. Similarly unsupported were News Plaintiffs' warnings that Copilot would destroy journalism. It turns out few users even use Copilot for news or current events—people are not cancelling newspaper subscriptions because they see Copilot as a substitute.

With every ability to opt out of web grounding, and no evidence of market substitution for their works, News Plaintiffs' claims would serve only to obstruct a powerful new tool for people to find and use information on the internet. News Plaintiffs evidently prefer traditional search engines, since they have built a business model around optimizing search results. But they have no right to the status quo. Copyright law exists to grow society's common intellectual wealth, not to lock in

entrenched business models or endow a select few outlets with a monopoly over sharing factual information.

This Court should grant summary judgment against News Plaintiffs' challenges to web grounding as well as to LLM training. Web grounding yields an immensely useful new tool that not only makes it easier for consumers to locate information, but also synthesizes information from numerous sources while tailoring a response to the user's request. It does so through intermediate copies that—by design and in actual practice—are not revealed to users. And this process provides answers that are not substitutes for news articles, in a product that is not a substitute for any news distribution platform. If website owners object anyway, they can opt out. It is hard to imagine a fairer use than one a copyright holder can block at will.

For similar but independent reasons, many of News Plaintiffs' claims fail because their knowing decision not to exclude their works from Copilot gives rise to an implied license. To afford website owners this prerogative, Microsoft developed and publicized Copilot specific "meta tags"—content controls widely used by entities like News Plaintiffs to control how their content is found and used. As courts have recognized in the context of traditional search, the knowing decision not to use controls like these constitutes implied consent to a location tool's use of that webpage.

Microsoft has deep respect for the critical role News Plaintiffs have long served in our public discourse, and Microsoft has partnered with news organizations to safeguard local and national journalism. But copyright law does not give News Plaintiffs the right to block transformative new technologies that do not replace their protected expression. The law expects rightsholders to adapt to technological progress, as they have successfully done before. The Court should grant summary judgment to Microsoft.

STATEMENT OF FACTS

Copilot pairs LLM technology with search infrastructure to create a powerful new knowledge and information tool.

In February 2023, Microsoft launched its first AI knowledge assistant, a chatbot built around OpenAI's GPT models. Canel ¶ 22. Before describing the product, a brief terminological aside is in order. Originally called Bing Chat, the product at issue in this case was renamed Copilot, a name it shares with many of Microsoft's AI-based offerings. The name Copilot is used in this brief for simplicity, and it refers to both the consumer-facing and enterprise assistant at issue in this case. The Court will recall a great many disputes over the scope of product discovery. This brief's description and depictions are based on record evidence concerning the versions of the product ruled in-bounds; the look and feel of those versions may differ slightly from newer versions of the product that are not at issue.

Though early AI chatbots were enormously powerful, they had an inherent limitation: Their generated outputs were frozen in time, based only on the parameters learned from their training data, and they could never learn more. SUMF ¶ 66. This left gaps, particularly on topics not covered by the training set. LLM-based chatbots can also engage in "hallucinations"—seemingly plausible but factually inaccurate responses. Abib ¶ 10.c.

Microsoft sought to overcome these limitations by pairing an LLM-based chatbot with the technological infrastructure underlying its Bing search engine. SUMF ¶¶ 67-71. This allows Copilot to retrieve relevant information from across the internet, then use the information to provide more accurate and contextually appropriate responses to user queries. SUMF ¶ 68. The marriage of the two technologies also offers a glimpse

into the next generation of web search. Rather than providing a static list of results, Copilot uses the LLM's language and reasoning to discern the user's intent, synthesize an answer from numerous sources of information, and field follow-up questions. SUMF ¶ 71; Abib ¶ 13.

With these capabilities, Copilot can tackle in seconds tasks that previously required hours of cumbersome research, plus the drudgery of culling and synthesizing information—whether it is planning a vacation, researching space exploration, choosing a college, fixing the plumbing, or countless others. Web grounding is what supplies the LLM powering Copilot with the best information to help with these tasks.

Web grounding uses retrieval augmented generation, or RAG, to ground responses in external webpages drawn from the Bing search index.

Here is how Copilot web grounding works. The technique Copilot uses to enable web grounding is called Retrieval Augmented Generation ("RAG"). SUMF ¶ 71. RAG is the set of technological processes that enables an LLM to retrieve materials from an external source, then ground the new, generated response from the LLM in the content that has been retrieved. SUMF ¶ 71. The external sources at issue here are webpages from across the internet, hence the term "web grounding."

When Copilot uses RAG technology to ground on potentially useful web content, it relies on Microsoft's Bing search index. SUMF ¶ 71.b. All conventional search engines, like Bing and Google, build search indexes with automated bots that continuously crawl every corner of the ever-changing internet, then analyze and organize the compiled information. Gaggara ¶¶ 7-8. The resulting index is a structured repository of links to webpages and stored webpage content itself, all of which allow the

search engine to near-instantaneously rank and retrieve results in response to a search. Gaggara ¶ 7.

Because it effectively acts as a front door to the internet, the index aims for broad coverage, Gaggara ¶ 18—and the fact that search engines are generally quite useful suggests most website owners appreciate the benefits of being findable online. But search norms have long given website owners the option to exclude their sites from search through "content controls"—costless, easily implemented technical directives applied to websites or particular pages. Canel ¶¶ 7-13. Microsoft has always honored content controls. Canel ¶ 8. It is well established that this standard approach to web indexing, including the copying of website content for the index, is fair use. *See Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003); 4 *Nimmer on Copyright* § 13F.14[B] ("wholesale duplication" of websites to create a search index is fair use).

Web grounding uses RAG to leverage Bing's existing search infrastructure. Think of web grounding as a series of pipelines, with information flowing between various technological components, all coordinated by an orchestration layer. SUMF ¶ 71; Abib ¶¶ 11-22. First, Copilot analyzes the user prompt to determine whether web grounding would improve the chatbot's response; if so, it optimizes the user's query for web search. SUMF ¶¶ 71.a-b. The optimized query then goes to Bing's search algorithm, which returns top results from the Bing Index. SUMF ¶ 71.b. For the top results, the orchestrator retrieves some portion of webpage content from the Bing Index and sends it to the LLM. SUMF ¶ 71.b-d.

Then comes the grounding step. As discussed in Microsoft's Books brief, an LLM's output is essentially a statistical prediction of successive tokens in a sequence based on some input (like a user's query) and the parameters the LLM learned from the

training data. With RAG, grounding content becomes *part of the input*. Content retrieved from the Bing Index is tokenized, and the tokens are placed in the LLM's "context window," alongside a tokenized version of the user's query. SUMF ¶ 71.d. These inputs then collectively influence the LLM's generated output by skewing the LLM's statistical prediction of the tokens that come next in sequence. SUMF ¶ 71.e. Grounding content is thus fundamentally *intermediate*—it is retrieved not to be deterministically displayed to the user, but instead to provide behind-the-scenes contextual information that affects the LLM's generated output. SUMF ¶¶ 72, 76-77.

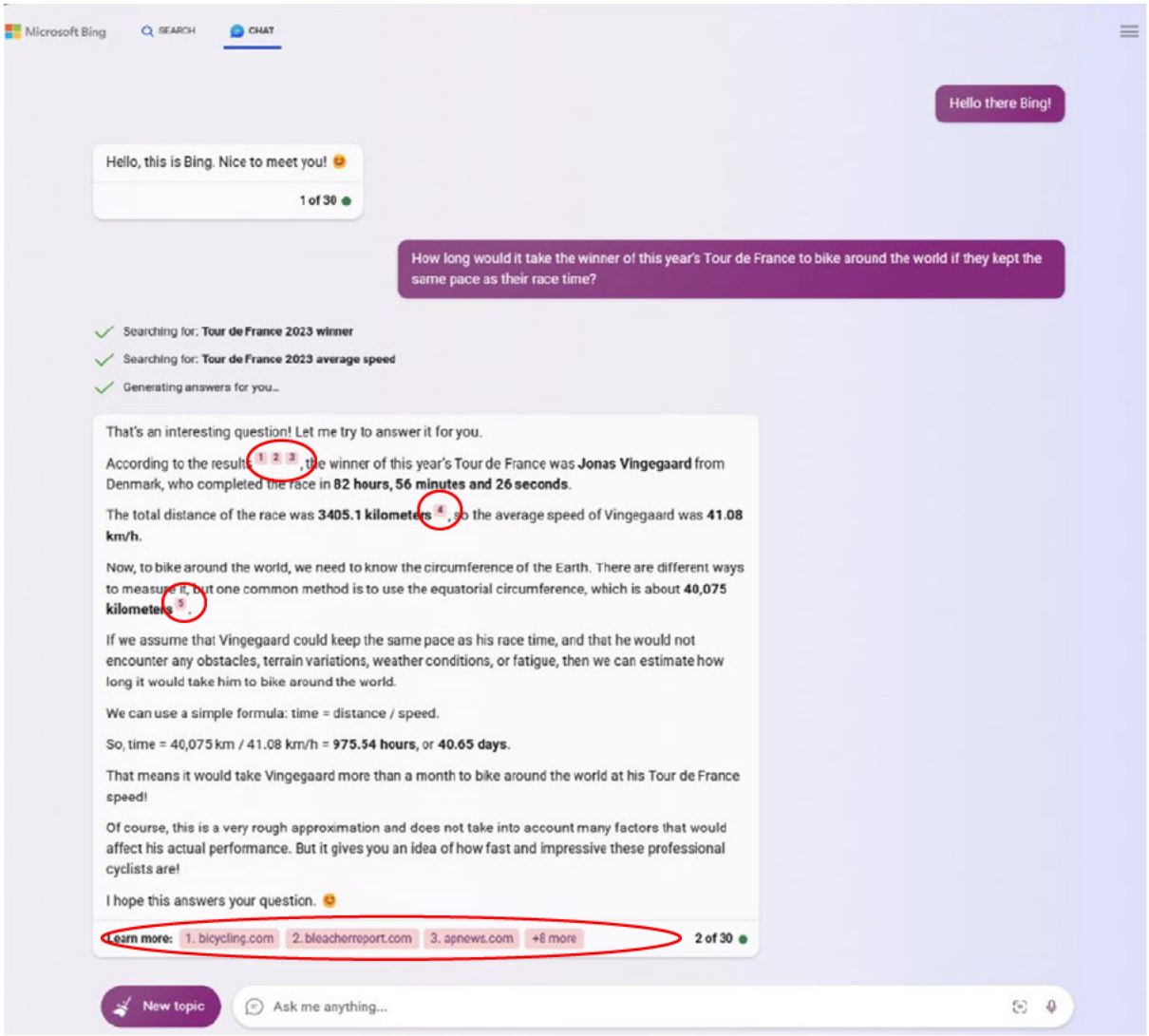
Microsoft designed Copilot in accord with its commitment to responsible AI.

Microsoft was one of the earliest large technology companies to develop responsible AI practices, choosing to formalize and operationalize those practices long before the current generative AI revolution. Bird ¶ 12. As LLM technology took off, Microsoft sought to develop and use the technology responsibly and to encourage others to do the same. Bird ¶¶ 16-20. It was important to Microsoft to strike an appropriate balance between the interests of rightsholders and the public interest in open access to information. Bird ¶¶ 43.c-d.

Though an LLM's non-deterministic prediction could incidentally include strings of tokens matching existing content, Copilot includes layers of mitigation to prevent display of copyrighted content to users. SUMF ¶ 74. Copilot has safeguards that recognize and refuse various types of user queries that request copyrighted works or portions of them. SUMF ¶ 74; Bird ¶ 30. Microsoft engineers also developed technological mitigations to prevent the LLM from even *incidentally* generating output that contains some non-trivial overlap with copyrighted content. SUMF ¶¶ 74-78. For

example, it used "metaprompt instructions" to direct the LLM not to generate outputs containing substantial verbatim overlap with grounding content. Bird ¶ 30.

Microsoft also carried forward two aspects of traditional search. The first was providing source attribution when Copilot grounds on a website, with links that allow users to navigate to the source website:



Hurst Ex. 184 at 4 (circles added).

The second is that Microsoft pioneered the adoption of content controls that enable any website owner to opt out of Copilot web grounding if they wish. Microsoft

believes that its products carry benefits for everyone in the internet ecosystem, but it is also important to be a collaborative partner with news publishers. Canel ¶¶ 23-25. In September 2023, Microsoft announced that website publishers could use a meta tag called "NOARCHIVE" to exclude web content from grounding by Copilot altogether, or another called "NOCACHE" to limit the amount of text used for grounding. SUMF ¶¶ 81-83. Publishers can also update meta tags for existing web pages, then use the "IndexNow" tool to have the webpage updated in the Bing Index to reflect the exclusion. Canel ¶ 19; Gaggara ¶ 27. And Microsoft's search team worked to address any concerns from rightsholders concerning their use of meta tags or technological glitches that arose. Canel ¶¶ 23-31.

News outlets claim that training LLMs and web grounding are copyright infringement.

Three groups of news organizations—(1) the New York Times ("the Times"), (2) the Daily News Plaintiffs, and (3) the Center for Investigative Reporting ("CIR")—sued Microsoft and OpenAI in 2023 and 2024. The Court dismissed several of News Plaintiffs' claims at the pleading stage. *See N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 297, 316-23 (S.D.N.Y. 2025) (claims under Digital Millennium Copyright Act and for "hot news" misappropriation). They voluntarily dismissed several more after discovery. ECF 1677 at 65-66 (NYT Compl.) (claims for contributory infringement and trademark dilution withdrawn); ECF 777 at 88 (No. 1:24-cv-03285) (NYDN Compl.) (same); ECF 1620 at 1 n.1 (CIR Compl.) (same for contributory infringement).

The operative Complaints assert direct and vicarious copyright infringement claims predicated on two alleged uses of their works: (1) OpenAI's use of news articles to train certain of its GPT models; and (2) alleged copies resulting from Copilot's web

grounding functionality. *See* NYT Compl. ¶¶ 168-71; NYDN Compl. ¶¶ 193-97; CIR Compl. ¶¶ 136-39. Microsoft seeks summary judgment against all remaining claims.

ARGUMENT

I. Training An LLM Is Fair Use As A Matter Of Law.

Microsoft's Books MSJ explains that the use of copyrighted content to train an LLM—from content acquisition to finished model—is fair use as a matter of law. *See* MSFT Books MSJ 10-33. The record in the News cases supports the same conclusion for much the same reasons. The Court should grant summary judgment against all claims of direct infringement against Microsoft related to model training. *See* NYT Compl. ¶¶ 165-75; NYDN Compl. ¶¶ 190-201; CIR Compl. ¶¶ 134-41.

A. Just like use of books to train an LLM, use of news content for training is highly transformative. *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1021-22 (N.D. Cal. 2025); *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1044-48 (N.D. Cal. 2025); MSFT Books MSJ 10-16. News articles are authored to be read by humans as expressive, journalistic works. SUMF ¶ 21. The "purpose and character," § 107(1), of using a news article to train an LLM is very different: to develop a technology that encodes human language and knowledge into statistical parameters capable of generating natural-language outputs. SUMF ¶¶ 1-3.

The transformative purpose and character of LLM training tilt every other factor towards fair use as well. MSFT Books MSJ 24, 25. The "nature" of the plaintiff's work has little role to play when the defendant's copying transforms the underlying work. *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 98 (2d Cir. 2014). If this factor did matter, it would weigh in favor of fair use here because News Plaintiffs' works are predominantly factual. SUMF ¶¶ 24-26.

As for the third factor, even "[c]omplete unchanged copying … [is] justified as fair use when the copying was reasonably appropriate to achieve the copier's transformative purpose." *Authors Guild v. Google, Inc.*, 804 F.3d 202, 221 (2d Cir. 2015). Here, every challenged use served to develop an LLM, and Microsoft's copying of entire works was "reasonably appropriate" to that purpose. SUMF ¶ 4.

And as to the fourth factor, economic "harm[s] [that] result[] from a transformative secondary use" carry no weight. *On Davis v. Gap, Inc.*, 246 F.3d 152, 175 (2d Cir. 2001). Factor four "is concerned with secondary uses that … usurp a market that properly belongs to the copyright-holder." *Infinity Broadcast Corp. v. Kirkwood*, 150 F.3d 104, 110 (2d Cir. 1998). Because transformative uses "do not serve as substitutes for the original work," *HathiTrust*, 755 F.3d at 99, they do not usurp the copyright holder's rightful market.

B. Most of News Plaintiffs' arguments against fair use are identical to those advanced by Books Plaintiffs, and fail for the same reasons. *See* MSFT Books MSJ 20-22 (memorization/regurgitation), 27-29 (claimed market for training data), 29-30 (value of works to training). We address three News-Plaintiffs-specific points below.

No substitution in traditional news markets. News Plaintiffs all earn the vast majority of their revenues from their works by one or more of (1) selling news content directly through subscriptions (paper or digital) or newsstands, (2) displaying advertisements in connection with their news content, (3) licensing content to other platforms, or (4) soliciting donations. SUMF ¶¶ 27, 33, 36, 39, 42, 45, 48, 51, 54, 57. Acquiring a news article, tokenizing it, and training an LLM on those tokens does not substitute for a purchase in any of those markets. Nor, as explained below when we address News Plaintiffs' challenge to web grounding (at 20-23, 23-28), is there evidence

that downstream uses of LLMs are generating substitutive articles that consumers can or do "acquire … in preference to the original," *Capitol Recs., LLC v. ReDigi Inc.*, 910 F.3d 649, 662 & n.16 (2d Cir. 2018) (citation omitted).

Market dilution harms are not cognizable. News Plaintiffs advance their version of Books Plaintiffs' "market dilution" theory—that LLM development could lead to creation of LLM-based tools that people use to create *non-infringing* news content that competes with News Plaintiffs' content. Several of News Plaintiffs' own witnesses do not believe this speculation. Hurst Ex. 123, 33:2-8 ("I do not view generative AI as a substitute for published news/advertorial environment content, no."); Hurst Ex. 11, 245:4-11 ("[Generative AI tools] are not a substitute for what journalists are capable of doing."). But even if dilution were a realistic threat, the market-dilution concept turns copyright upside-down. Competing non-infringing works are copyright goods, not evils. MSFT Books MSJ 30-32.

News Plaintiffs get no further by raising the specter of so-called "pink slime" journalism, Hurst Ex. 142 at 16—what their expert Emily Bell defines as "low quality algorithmically generated websites" masquerading as "authentic news sites." Hurst Ex. 228 ¶¶ 128. Bell acknowledges that pink slime existed before generative AI, but warns of "the possibility" that Defendants' LLM technology "would exacerbate the problem." Hurst Ex. 228 ¶ 133. If "pink slime" websites are publishing "plagiarized articles … stole[n] word-for-word" from News Plaintiffs, Hurst Ex. 228 ¶ 136, that is terrible. But there is no evidence those articles are coming from use of Defendants' LLMs. News Plaintiffs should sue the pink slime websites.

If instead News Plaintiffs' complaint is that Defendants' LLMs could be built into products that users then utilize to author *non-infringing* works, the argument fails just

the same. News Plaintiffs have no right to prohibit others from reporting facts. *N.Y. Mercantile Exch. v. IntercontinentalExchange, Inc.*, 497 F.3d 109, 114 (2d Cir. 2007). Nor do they have any right to prevent market harms flowing from works that do not exploit their protected expression. *See* MSFT Books MSJ 30-31.

Post-training uses to develop an LLM are fair. For one model at issue in the News (but not Books) cases, called Deucalion, Microsoft engaged in "post-training" conduct. Part of this post-training involved creating and showing the model synthetic data derived from user conversations, then providing the model feedback on its responses to improve performance. SUMF ¶¶ 17-18. During discovery, News Plaintiffs noted snippets of their webpages in 45 such synthetic conversations out of 228,679 total messages. Lafferty ¶¶ 336-39. These minor uses are fair use as a matter of law for the same reason most any use in developing an LLM is fair, so any attempt to predicate a separate claim against Microsoft on this conduct fails.

***

Allowing News Plaintiffs or other copyright owners to demand a license for use of works to train an LLM would "stifle innovation" and ultimately "the very creativity which [copyright] is designed to foster." *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 18, 21 (2021). This Court should hold that the use of news content to train an LLM is fair use, entitling Microsoft to summary judgment on all direct and secondary infringement claims arising from LLM training.

II. Copilot Web Grounding Is Fair Use As A Matter Of Law.

News Plaintiffs' remaining claims against Microsoft challenge Copilot's web grounding. Plaintiffs advance two theories. The first focuses on the *inputs* to the process—the grounding content copied when, in response to a user query, Copilot

searches the Bing Index, retrieves a webpage, and grounds the LLM's response on that content. The second theory targets *outputs*: News Plaintiffs assert that if Copilot-generated responses contain non-de-minimis verbatim overlap with their works, those outputs infringe. (Spoiler: they virtually never do.)

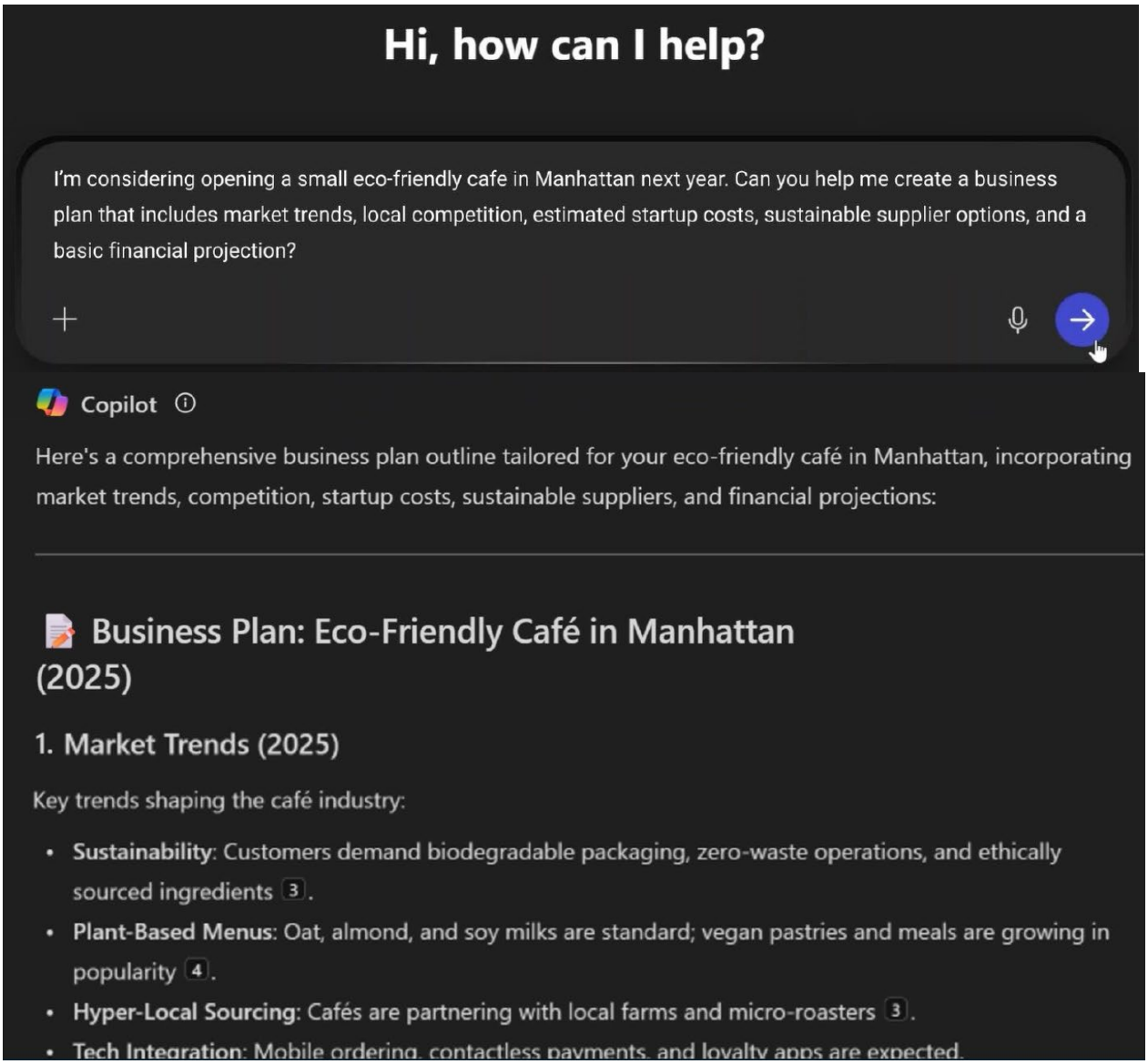
Both theories fail. Web grounding creates a transformative new product much different from News Plaintiffs' works, *infra* § A; relies on intermediate copies that are not revealed as outputs to the user, *infra* § B; and does not substitute for News Plaintiffs' expressive content in any cognizable market, *infra* § C. This Court should therefore hold that web grounding—including both grounding copies used as inputs and any rare, incidentally generated outputs that match short strings of grounding content—is fair use. NYT Compl. ¶¶ 165-75; NYDN Compl. ¶¶ 190-201; CIR Compl. ¶¶ 126-33.

A. Copilot web grounding creates a new product for finding and using information (Factor one).

As explained previously, whether a use of copyrighted works is transformative lies at the heart of the fair use analysis, bearing on every factor. *Supra* 10; MSFT Books MSJ 10, 24, 25. The purpose of news articles is to provide directly sourced, timely journalistic reporting on events or other topics selected by the publication and author for their readership. SUMF ¶ 21. Copilot web grounding serves new and different purposes. It enables Copilot's LLM to generate more accurate, informed, and direct responses to a broad range of user queries. The result is a transformative "new product," *Oracle*, 593 U.S. at 30, that in no way resembles newspapers, news websites, news aggregators, or any other online distributor of journalism.

As explained above (at 4), an LLM response without some external source of information will be based solely on the LLM parameters calculated during training.

SUMF ¶ 66. So, like a human answering based on background knowledge and fuzzy memory, an ungrounded LLM can struggle with precision or guess to fill in gaps. Web grounding dramatically improves LLMs by giving them access to a vast, continuously updated external source of information: the internet. SUMF ¶¶ 67-68. The internet knows about the latest deadline for casting absentee ballots in your state, how K-Pop evolved into a global phenomenon, and where to get the best brisket in Texas Hill Country—and it knows such things based on many types of webpages, including, but certainly not limited to, news articles. Grounding responses on search results gives users fresher, more useful responses. SUMF ¶¶ 67-68. And by providing links to grounding sources, Copilot allows users to verify the reliability of responses and learn more. SUMF ¶ 73.



ECF 274 at 46-47 (showing full output).

As the above example illustrates, web grounding also does more than just enhance the "use and usefulness" of AI chatbots. *Oracle*, 593 U.S. at 30. It supplies powerful functionality beyond basic web search (itself a highly valuable and transformative use). *See Authors Guild*, 804 F.3d at 217 ("add[ing] important value to the basic transformative search function" is transformative). Traditional search engines are excellent information location tools, and still often the quickest way to find a

particular thing online. But a web grounded LLM can handle far more complex tasks with ease. *Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169, 177 (2d Cir. 2018) ("enhanc[ed] efficiency" is itself a "transformative purpose").

To begin with, an LLM's language processing capabilities are better able to discern what a user query is really after and therefore retrieve the best results. Abib ¶¶ 6-7, 22; Hurst Ex. 294, 112:24-113:2. The LLM can synthesize information from within web grounding results in light of the user's (LLM-optimized) query. SUMF ¶ 72. And the chatbot interface allows users to iterate a query over time by asking follow-up questions or "engag[ing] in a dialogue" with Copilot. Abib ¶ 6. For a user, the experience is like having a conversation with the internet about what it knows, how it knows it, and where to go to learn more. Copilot thus does many things a traditional search engine cannot do, or does them much more efficiently.

These purposes—increasing accuracy, timeliness, relevance, and usefulness—entail something new and different than the underlying grounding content. Whether results used for web grounding include news articles or some other type of content, that content is not revealed to the user, but instead tokenized and placed in the LLM's context window to influence the response. SUMF ¶¶ 71.d, 77. The resulting response is a novel, generated synthesis of numerous sources, amalgamated with and by the LLM's existing parameters—adding further "meaning" and "message" that differs from any individual piece of grounding content. *Oracle*, 593 U.S. at 29 (citation omitted); SUMF ¶ 72. This is so whether Copilot is grounding on a news article or anything else.

Copilot is thus not a product for distributing news articles or any other copyrighted content. It is more like a "multifaceted research service," *Swatch Grp. Mgmt. Servs Ltd. v. Bloomberg L.P.*, 756 F.3d 73, 83 (2d Cir. 2014), that uses

intermediate sources to "furnish[] … valuable information on any subject of public interest … without allowing public access to the copy," *Romanova v. Amilus Inc.*, 138 F.4th 104, 109, 115-17 (2d Cir. 2025). This transformative purpose means that it matters little that Copilot is a product of a "commercial nature," § 107(1). *Consumers Union of U.S., Inc. v. Gen. Signal Corp.*, 724 F.2d 1044, 1049 (2d Cir. 1983). And it matters even less in light of Copilot meta tags, which allow any website owner to exclude their content if they wish. SUMF ¶¶ 81-83. This further underscores that the purpose of web grounding is not to hawk others' expression for profit, but to help users locate information in order to "enrich[] public knowledge." *Authors Guild*, 804 F.3d at 214. Factor one weighs heavily in favor of fair use.

B. Intermediate grounding copies use factual elements and do not display protected expression to users (Factors two and three).

The "nature of the copyrighted work," § 107(2), and the "amount and substantiality … used," § 107(3), also weigh in favor of fair use.

1. News articles are predominantly factual.

News Plaintiffs assert "predominantly factual news articles." *Nihon Keizai Shimbun, Inc. v. Comline Bus. Data, Inc.*, 166 F.3d 65, 72 (2d Cir. 1999); *see* SUMF ¶¶ 24-26. "It is well established that 'the scope of fair use is greater with respect to factual rather than non-factual works,'" *Swatch*, 756 F.3d at 89 (citation omitted), because the aims of copyright are not advanced by "restraining the free flow of information." *Consumers Union*, 724 F.2d at 1049. Any member of the public has "every right" to "republish the facts contained in [News Plaintiffs'] articles." *Nihon*, 166 F.3d at 70; *see Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 348 (1991) (the "news of the day … may not be copyrighted and [is] part of the public domain available

to every person"). News Plaintiffs can hardly claim otherwise—every day they republish news first reported by others. *See* Hurst Ex. 24, 93:25-96:16; Hurst Ex. 93, 74:23-75:11. Factor two thus weighs in favor of fair use.

2. Grounding copies are intermediate versions made for purposes of influencing generated responses.

As noted, News Plaintiffs challenge both use of their content for grounding and alleged display of their works in outputs. Because the "amount and substantiality" of the works used differ considerably between grounding and outputs, we address the two in turn.

In Copilot, grounding copies are tokenized versions of webpages (or portions of webpages) that act as temporary inputs to influence the LLM's response. SUMF ¶¶ 71.c-d, 77. Although all tokens from a news article could be used for grounding, their purpose is to provide Copilot's LLM with raw material that enables generation of accurate, contextualized results. SUMF ¶ 77; *see Authors Guild*, 804 F.3d at 220 (even "[c]omplete unchanged copying" is fair use if "reasonably appropriate" to transformative purpose).

As explained above (at 6-7), Copilot's responses are generated anew from the LLM, regardless of the amount of the underlying news article (or any other source) the LLM consults. There is no deterministic functionality that passes content through to the user. SUMF ¶ 76. Again, as intermediate copies, grounding copies exist inside Copilot's retrieval pipeline, enabling the LLM to generate output according to the LLM's parameters, as influenced by the contents of the context window. *Supra* 6-7.

And in all events, website owners can *control* the "amount and substantiality" of the grounding copy used through Microsoft's meta tags. Use of NOARCHIVE excludes a

webpage from grounding entirely. SUMF ¶ 82. And use of NOCACHE limits grounding to only a portion of webpage. SUMF ¶ 82; *see supra* 8-9, *infra* 33 (discussing meta tag functionality). The amount used for grounding is therefore only so much as the website owner allows.

In short, grounding copies are the sort of intermediate copies never "made accessible to" the public that courts have repeatedly found to be fair use. *Authors Guild*, 804 F.3d at 222 (emphasis omitted); *Sony Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 605-06 (9th Cir. 2000); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1522-24 (9th Cir. 1992); MSFT Books MSJ 14-15 (discussing cases).

3. Copilot responses displaying incidental verbatim fragments of News Plaintiffs' works are exceptionally rare.

As for News Plaintiffs' fears that Copilot would *output* passages of their works to users, discovery proved them unfounded. Copilot doesn't do that. At worst, this record shows exceptionally rare instances of Copilot incidentally generating fragmentary matches with News Plaintiffs' works. SUMF ¶ 79. Even then, the outputs in the record often contain overlap consisting of mere factual details and basic ideas. If any of this even rises to the level of an actionable copyright event—and that is doubtful—these minor uses are fair for the same reasons as web grounding generally.

To start, it is undisputed that Microsoft developed numerous measures designed to drive matching outputs "as close to zero as possible," Hurst Ex. 294, 107:12-109:4—including a highly effective, standing "metaprompt" *instructing* the LLM "to not generate outputs containing substantial verbatim overlap," Bird ¶¶ 38-39; *supra* 7-8. Absolute zero is difficult to achieve, because sometimes an LLM influenced by web grounding may incidentally predict a string of tokens that matches fragments of

grounding content, or it may attempt (without mitigations) to satisfy a user query by providing strings of tokens. SUMF ¶ 78. But absolute zero is also unnecessary from a copyright perspective. "The New York Knicks won the NBA Championship for the first time in 53 years" is a 15-word factual statement surely recited all over the internet. And rarely does a news article itself *not* contain a quotation that is similar in length. Repeating these things is not a copyright problem. Courts have recognized that a mere "one sentence" of a literary work is "ordinarily" considered *de minimis* and therefore "not actionable." *See Hanagami v. Epic Games, Inc.*, 85 F.4th 931, 946 (9th Cir. 2023). Yet still, over 99% of Copilot outputs contain not even that minimal degree of verbatim overlap with News Plaintiffs' works. SUMF ¶ 79.

News Plaintiffs have spent this case hypothesizing that users could nevertheless find ways to dupe Copilot into outputting significant portions of works. Summary judgment is no time for hypotheses, nor for "unfounded speculation" based on one-off examples of mischievous users trying to Rube-Goldberg content out of a chatbot. *See A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 645 (4th Cir. 2009) ("scenarios envisioned by plaintiffs" were "implausible in light of how the [challenged] system generally operates"); *Authors Guild*, 804 F.3d at 222 (plaintiffs' theory that copyrighted content was "*theoretically* accessible" did not prove it was "in fact accessible"). What is material now is how Copilot actually behaves with respect to News Plaintiffs' works.

Here, discovery brought News Plaintiffs what ought to be good news: Copilot does not produce meaningful displays of Plaintiffs' content. To evaluate this, Microsoft provided an agreed-upon sample of 8.2 million chat logs containing user prompts, grounding content, and Copilot responses. The 8.2 million were *not* random—they were

the chats requested by Plaintiffs because they hit on keywords implicating use of News Plaintiffs' websites, and therefore the *most* likely to contain News Plaintiffs' works. Lafferty ¶ 27. The parties' experts scrutinized these logs in various ways, and for present purposes, we will put methodological disputes to the side because they all converge on the same conclusion.

Tom Goldstein, expert for the Times and Daily News, filtered the 8.2 million chat logs for conversations in which Plaintiffs' domains appeared in the content that Copilot used for grounding. SUMF ¶ 79. He then searched for 16-word matches between the grounding content and the Copilot response. SUMF ¶ 79. He could find only 59,545 conversations that had matching content even the length of this sentence. SUMF ¶ 79. That is itself far shorter than many a quotation contained in and repeated across news articles. As Microsoft's expert calculated, dividing Goldstein's number by the 8.2 million chat logs produced yields a miniscule 0.73% rate of matches, even at Goldstein's de minimis threshold. Lafferty ¶ 399. As for CIR's expert, Emily Wenger, she found even tinier rates of overlap with CIR works in Copilot's 8.2 million outputs; she found only *51* instances of her own definition of what constitutes substantial overlap. SUMF ¶ 79.

Even crediting News Plaintiffs' experts, the most they establish is that, in exceptionally rare cases, outputs grounded on News Plaintiffs' works could contain small strings of generated text incidentally matching News Plaintiffs' works. Because the third factor turns on the "amount and substantiality" of copied works "*made accessible* to a public for which [they] may serve as a competing substitute," *Authors Guild*, 804 F.3d at 222, that factor weighs strongly in favor of fair use here. *See id.* (deliberate snippet view feature revealing three lines of a book is fair use); *Ringgold v.*

Black Ent. Television, Inc., 126 F.3d 70, 75 & n.4 (2d Cir. 1997) (noting relationship between de minimis copying doctrine and fair use).

C. Web grounding is not a substitute for News Plaintiffs' works in any cognizable market (Factor four).

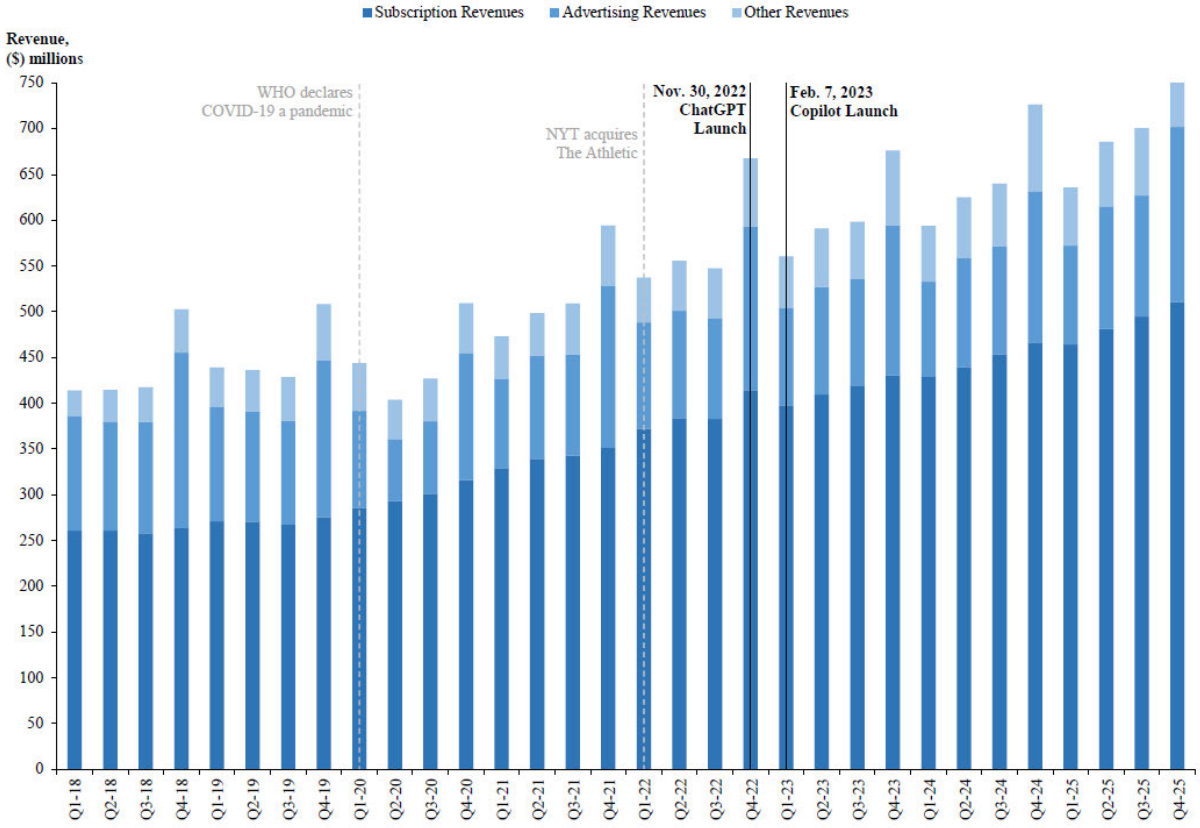
As a transformative use that reveals virtually no protected expression to users, web grounding does not harm any market "properly belong[ing]" to News Plaintiffs. *Infinity Broadcast*, 150 F.3d at 110; § 107(4).

1. Plaintiffs cannot demonstrate harm in any traditional market.

Copilot has had no effect on "any part of the normal market" for News Plaintiffs' works. *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 568 (1985) (internal quotation omitted). Because Copilot does not output substantial copies of articles, one cannot "acquire the copy in preference to the original." *Authors Guild*, 804 F.3d at 223. The economic evidence proves it beyond doubt. News Plaintiffs' revenues, subscriptions, donations, web traffic and conversion rates have either increased, remained stable, or trended in the same direction at the same rate in the period both before and after the launch of Copilot (and ChatGPT). SUMF ¶¶ 85-144.

Microsoft's economic expert, Catherine Tucker, presented all of this data in detail in her reports. Tucker Decl. ¶¶ 11(a), 16-95, 106-116. Take, for example, the Times, which just publicly released blockbuster Q2 growth while touting "the durability of our business model." Hurst Ex. 262 at 1.

Undisputed evidence shows that revenues are up since Copilot's launch.



Subscriptions are up, too.



Website traffic has been stable or up:



And conversion rates—i.e., the rate at which a website user signs up for a subscription after being served a paywall, Tucker ¶ 52—are undiminished.



For every News Plaintiff, the data shows that the launch of LLM-based chatbots, and web-grounded Copilot in particular, had no effect on the direction or rate of their metrics. *See* Tucker ¶¶ 11(a), 16-95, 106-116.

News Plaintiffs have resisted this conclusion by suggesting that Copilot competes with their works for clicks and eyeballs in the broader attention economy. *See, e.g.*, Hurst Ex. 11, 52:1-8. But this attention-market theory fails because copyright law gives News Plaintiffs no *right* to attention. They have the right only to prevent exploitation of their creative expression. Fair use is designed to *promote* transformative uses that consumers value, not ban them.

For similar reasons, it would do News Plaintiffs no good to posit that their market extends so broadly as to cover any demand for news or information. Any such market does not "properly belong[]" to News Plaintiffs. *Infinity Broadcast*, 150 F.3d at 110; *supra* 11. In *Authors Guild*, for example, the Second Circuit recognized that a Google Books "snippet" might "satisfy a searcher's need for … [the] book" from which it is excerpted if the snippet suffices to "convey[] a … fact that the searcher needs to ascertain." 804 F.3d at 224. But the court held that this is not "the type of loss of sale" copyright protects against, because copyright "protects only the author's manner of expression." *Id*. Likewise here. News Plaintiffs are entitled to no monopoly over answering questions about the who/what/when/where/why/how of human affairs.

In any event, Copilot usage data demonstrates that News Plaintiffs' fears that Copilot is a substitute source of journalism are speculative. Microsoft's expert evaluated a representative sample of Copilot conversations and found that just 1.3% of end-users use Copilot to research current events at all. SUMF ¶ 145. An additional survey of users confirmed Copilot's low rate of usage for news. SUMF ¶ 145. Copilot is not a source of journalism; it is a "very different type of product" operating in a "distinct" market. *Oracle*, 593 U.S. at 37 (citation modified).

Three years of economic evidence proves that Copilot web grounding is not a substitute for news articles. Factor four therefore weighs in favor of fair use.

2. The Court should not recognize a market for grounding content.

News Plaintiffs also claim harm from lost licensing revenue they would have received had Microsoft paid for web grounding. *See, e.g.*, Hurst Ex. 136 at 7; Hurst Ex. 154 at 8-9. If News Plaintiffs are claiming a market for *intermediate grounding copies*,

the argument fails for the same reason Plaintiffs' argument about a market for *LLM training data* fails. *Supra* 11; MSFT Books MSJ 27-28. Undisputed evidence establishes that no such market existed or was likely to develop prior to the advent of web grounding itself. SUMF ¶¶ 147, 149. Courts have rejected plaintiffs' reliance on a "potential market" for the very transformative use they are challenging. *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 929 n.17 (2d Cir. 1994) (citation omitted). And they have also rejected reliance on scattered licensing deals for the challenged use, explaining that "copyright owners may not preempt exploitation of transformative markets … by actually developing or licensing others to develop those markets." *Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc.*, 150 F.3d 132, 145 n.11 (2d Cir. 1998). *See* MSFT Books MSJ 28.

If News Plaintiffs are instead suggesting that they are losing licensing revenue in the market for *display* of their articles, the answer there is simple: Copilot does not display their articles. There is no market for rare, de minimis generation of fragments of text, which is all News Plaintiffs can show. *Supra* 20-23. So insofar as News Plaintiffs point to deals that include *display* rights—even deals Microsoft has entered to explore possible display in generative AI products—those deals are irrelevant to the product at issue here. *See generally* Tucker ¶¶ 177-79 (explaining economic irrelevance of licenses for display rights).

3. News Plaintiffs are not entitled to entrench their business models.

News Plaintiffs' last theory of economic harm posits that web grounding could be a substitute for traditional search engines—a puzzling argument, since none of them run a search engine. Hurst Ex. 142 at 17-18; Ex. 154 at 6-7. News Plaintiffs, the theory goes,

have learned to "strategically position[] themselves [within]" traditional search engines to "drive[] … user engagement" (i.e., clicks and traffic to their sites). Hurst Ex. 244 ¶ 128. They say they cannot do this as effectively with generative AI tools, Hurst Ex. 244 ¶¶ 332, 386, 389, pointing to an expert's calculation that users "click through" to their websites at a lower rate from Copilot than they do from the traditional Bing search engine. They posit that this will lead to reduced traffic to their webpages, in the face of three years of evidence saying it will not, *supra* 23-27.

This argument fails as a matter of law and fact. The legal problem with News Plaintiffs' theory is that it claims not *direct* harm from a *substitutive* use, but *indirect* harm from a *transformative* one. Only the former counts under factor four. As *Oracle* explains, it may well be that a transformative use undermines a copyright holder's business model and thereby hampers "demand for the original." 593 U.S. at 35 (citation omitted). But absent substitution, such harms are not "cognizable" because a copyright owner has no right to "lock" in an advantageous market dynamic. *Id.* at 35, 39 (citation omitted). Such "illegitimate" use of copyright protection would "interfere with, not further, copyright's basic creativity objectives." *Id.* at 22, 39.

News Plaintiffs' theory suffers from precisely this problem. Copilot web grounding is a transformative use that gives consumers a new product. If consumers prefer this new product to traditional search, and if that preference disrupts entrenched business models, that is not a copyright problem. News Plaintiffs are not entitled to lock in existing sources of referral traffic or historical click-through rates.

News Plaintiffs have failed to prove the theory anyway. For the click-through-replacement theory to work, News Plaintiffs would need to show (1) that users are indeed using Copilot *instead of*, and not *in addition to*, traditional search; and (2) that

the use of Copilot instead of search engines resulted in less traffic to their webpages. News Plaintiffs have failed to show either. On (1), their expert conceded that she had no opinion "about whether chatbots are replacing search engine behavior at scale." Hurst Ex. 114, 96:15-97:9. And on (2), again, neither total traffic nor conversions to News Plaintiffs' sites have changed as a result of Copilot. *Supra* 26-27. There is no evidence that Copilot has caused a decrease in traffic to News Plaintiffs' sites—in fact, the Times's and Daily News Plaintiffs' experts did not even analyze web traffic, despite presenting a theory of harm premised entirely on how Copilot could theoretically have an adverse impact on traffic. SUMF ¶¶ 88-91, 95-97, 101-103, 107-109, 113-115, 119-121, 125-127, 131-133 137-139, 143-146.

Once again, if News Plaintiffs disagree, they have recourse: They can use meta tags to exclude their content from Copilot, while keeping it in traditional search. But they may not prevent the rest of the world from moving forward. The Court should hold that Copilot web grounding is fair use as a matter of law and grant summary judgment.

III. News Plaintiffs' Knowing Decision To Forego Use Of Copilot Meta Tags Gives Rise To An Implied License.

Many of News Plaintiffs' grounding claims fail for the independent reason that News Plaintiffs impliedly licensed Copilot's use by knowingly declining to implement meta tags on their webpages. Summary judgment is warranted for any claim based on webpages lacking a NOARCHIVE tag after Microsoft announced this content control in September 2023. *See* SUMF ¶¶ 150-51.

A license is a defense to copyright infringement. *Yamashita v. Scholastic Inc.*, 936 F.3d 98, 104 (2d Cir. 2019). A nonexclusive license for a use may be implied by conduct when facts and circumstances indicate a "meeting of the minds." *ABKCO*

Music, Inc. v. Sagan, 50 F.4th 309, 320 (2d Cir. 2022) (citation omitted); *see* 3 *Nimmer on Copyright* § 10.03[A]. In two influential decisions, district courts have evaluated implied licenses in the context of internet search. *See Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1115-16 (D. Nev. 2006); *Parker v. Yahoo! Inc.*, No. 07-2757, 2008 WL 4410095, at *4 (E.D. Pa. Sept. 25, 2008). These decisions turn on a well-known system of content controls specific to the internet.

As explained above (at 6), website owners have long used content controls to restrict whether and for what purposes "web crawlers" or "robots" can use their webpages. Gaggara ¶¶ 10-16; Canel ¶¶ 8-13; *see generally Field*, 412 F. Supp. 2d at 1110-11. The most commonly known is "robots.txt," a file that specifies to crawlers which pages on a website may be crawled. Canel ¶ 10. Meta tags provide more granular control—dictating, for example, that a webpage may be included in a search index but that no portion may be displayed as part of a search result. Canel ¶ 11; *see Field*, 412 F. Supp. 2d at 1112-13.

In *Field*, the plaintiff alleged that Google infringed by displaying an entire indexed (or "cached") copy of a webpage containing his copyrighted works. 412 F. Supp. 2d at 1115. But the plaintiff "knew that the presence of a 'no archive' meta-tag on the pages of his Web site would have informed Google not to" make this use. *Id.* at 1116. The court held that "cho[osing] not to include the no-archive meta-tag on the pages of his site …. with knowledge of how Google would use the copyrighted works he placed on those pages" gave rise to an implied license. *Id.* Similarly, *Parker* related to search engines' display of "unauthorized 'cached' copies" of the plaintiff's works. 2008 WL 4410095, at *1. The court found an implied license for the use because the plaintiff "knew" that the defendant search engines "honor[ed] 'electronic protocols' that would

prevent" it—namely, the "'robots.txt' protocol"—yet he "failed to employ" the exclusion. *Id.* at *4.

The same analysis applies here. Microsoft's Copilot-specific NOARCHIVE tag, which Microsoft publicized in September 2023, allows website owners to indicate that a webpage copied into the Bing Index can appear in *traditional* Bing search results, but *not* be used as grounding content for Copilot. SUMF ¶¶ 82, 154; Gaggara ¶ 15. News Plaintiffs indisputably knew about that tag, knew how to use it, and knew that if they did not, their content would be used for Copilot web grounding. SUMF ¶ 154. Several of News Plaintiffs' corporate representatives conceded as much directly. For example, Sam Felix, the Times's former Vice President, testified that the Times knew NOARCHIVE would disable use of their content for web grounding "once Microsoft told" the Times in September 2023. Hurst Ex. 34A, 117:10-118:23. And several News Plaintiffs have used the meta tags at times, confirming their knowledge of them. SUMF ¶¶ 155-58.

The Court should find that use of un-tagged web content after September 2023 is impliedly licensed for web grounding. Use of meta tags is simple, costless, and understood by webmasters like those employed by News Plaintiffs. Canel ¶ 10; SUMF ¶¶ 152-54. So there is no possible inference that News Plaintiffs elected not to use content controls because they were too hard or expensive to employ. Copilot meta tags, moreover, are narrowly tailored to web grounding. NOARCHIVE excludes content entirely from Copilot, but not traditional Bing search. A website owner's visibility through *traditional* search is thus not compromised or downgraded by their decision to exclude their content from Copilot.

This distinguishes the only case that has rejected an implied license based on non-use of internet content controls, *Associated Press v. Meltwater U.S. Holdings, Inc.*,

931 F. Supp. 2d 537 (S.D.N.Y. 2013). The problem there, the court found, was that mere non-use of a content control typically used *for search features* was not also implicit authorization for publication of the works in the distinct context of a daily newsletter. *Id*. at 563. Non-use of a content control needs to fairly imply "which *types* of use" are permitted. *Id*.

What was missing from *Meltwater* is present here. News Plaintiffs knew that declining to implement a NOARCHIVE meta tag on their works after September 2023 would permit exactly the use that they complain of here: use of those works as grounding content for Copilot. News Plaintiffs cannot complain that any use of those works for grounding was unauthorized.

IV. At Most, News Plaintiffs Are Entitled To One Award Of Statutory Damages Per Issue.

News Plaintiffs' claims against Microsoft should not move forward. But if they do, Plaintiffs' astronomical statutory damages demand needs some grounding of its own—in the plain text of the Copyright Act. Section 504(c)(1) explains that "[f]or the purposes" of statutory damages, "all the parts of a compilation … constitute one work." § 504(c)(1). A periodical containing news articles is unquestionably a compilation. *See* § 101 ("periodical issue" is a "collective work," and the "term 'compilation' includes 'collective works'"). So News Plaintiffs are entitled at most to one award per *issue*, not per *article*.

There is no reasonable dispute on this record that virtually all of News Plaintiffs' articles are parts of issues. Indeed, their uniform practice when they register their works is to register not individual articles, but *issues* as collective works. SUMF ¶ 160. That confirms that these articles are parts of compilations. And "the plain language of

the Copyright Act limits the copyright holders' statutory damage award to one for each" compilation. *Bryant v. Media Right Prods., Inc.*, 603 F.3d 135, 141 (2d Cir. 2010). Plaintiffs cannot multiply their damages by now recharacterizing each article as a standalone "work."

News Plaintiffs appear poised to make just this argument by relying on *EMI Christian Music Group, Inc. v. MP3tunes, LLC*, 844 F.3d 79 (2d Cir. 2016), which permitted record companies to recover individual awards for independently released singles, even those contained in albums. *Id.* at 101. But musical singles have long been distinct commercial works in the marketplace, with consumers purchasing them separately and radio stations playing the hits. No court has ever extended *EMI Christian* outside of that context, and News Plaintiffs cannot justify such an extension here.

Indeed, as evidenced by their registrations, the asserted articles were overwhelmingly released and published as part of newspaper issues, not as individual articles for sale to the public. SUMF ¶ 160. Nor can Plaintiffs simply state a willingness to sell or license articles individually—if that tactic worked, *EMI Christian*'s separate issuance test would swallow § 504(c)(1)'s compilation limitation entirely. If this Court declines to grant summary judgment for Microsoft on all claims, it should hold that any eventual statutory damages award must be calculated on a per-issue—rather than per-article—basis.

CONCLUSION

The Court should grant summary judgment in favor of Microsoft on all copyright claims challenging LLM training and web grounding, or at a minimum limit statutory damages to one award per issue.

Dated: September 4, 2026

Respectfully submitted,

/s/ Annette L. Hurst

ORRICK, HERRINGTON & SUTCLIFFE LLP
Annette L. Hurst (admitted pro hac vice)
Daniel Justice (admitted pro hac vice)
Robert Scott Shwarts
Catherine Lui (admitted pro hac vice)
Amanda H. Schwartz (admitted pro hac vice)
Ana Mendez-Villamil (admitted pro hac vice)
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669
Telephone: (415) 773-5700
Facsimile: (415) 773-5759
ahurst@orrick.com
djustice@orrick.com
aschwartz@orrick.com
amendez-villamil@orrick.com

Christopher J. Cariello
Lisa T. Simpson
Marc Shapiro
Vanessa Sorrentino
Hilda Obeng
51 West 52nd Street
New York, NY 10019
Telephone: (212) 506-3778
Facsimile: (212) 506-5151
ccariello@orrick.com
lsimpson@orrick.com
mrshapiro@orrick.com
vsorrentino@orrick.com
hobeng@orrick.com

Sheryl Koval Garko (admitted pro hac vice)
Laura Brooks Najemy (admitted pro hac vice)
222 Berkeley Street
Boston, MA 02116
Telephone: (617) 880-1800
Facsimile: (617) 880-1801
sgarko@orrick.com
lnajemy@orrick.com

Isaac Sawlih Behnawa (admitted pro hac vice)
Johnathan Vaknin (admitted pro hac vice)
355 S. Grand Ave. Suite 2700
Los Angeles, CA 90071
Telephone: (213) 612-2179
Facsimile: (213) 612-2499
ibehnawa@orrick.com
jvaknin@orrick.com

Monica Svetoslavov (admitted pro hac vice)
Louis Andrew Schreiber
2100 Pennsylvania Avenue NW
Washington, DC 20037
Telephone: (202) 339-8400
Facsimile: (202) 339-8500
msvetoslavov@orrick.com
aschreiber@orrick.com

FAEGRE DRINKER BIDDLE & REATH LLP
Jeffrey S. Jacobson
1177 Avenue of the Americas
New York, New York 10036
Telephone: (212) 248-3191
jeffrey.jacobson@faegredrinker.com

Jared B. Briant (admitted pro hac vice)
Kirstin L. Stoll-DeBell (admitted pro hac vice)
Timothy Scull (admitted pro hac vice)
Shelby Pickar-Dennis (admitted pro hac vice)
1144 15th Street, Suite 3400
Denver, Colorado 80202
Telephone: (303) 607-3588
jared.briant@faegredrinker.com
kirstin.stolldebell@faegredrinker.com
timothy.scull@faegredrinker.com
shelby.pickar-dennis@faegredrinker.com

Lora A. Brzezynski (admitted pro hac vice)
Brianna Lynn Silverstein (admitted pro hac vice)
Faith Maggard (admitted pro hac vice)
1500 K Street NW, Suite 1100
Washington, DC 20005
Telephone: (202) 230-5000
Facsimile: (202) 842-8465
lora.brzezynski@faegredrinker.com

brianna.silverstein@faegredrinker.com
faith.maggard@faegredrinker.com

Carrie A. Beyer (admitted pro hac vice)
320 South Canal Street, Suite 3300
Chicago, IL 60606-5707
Telephone: (312) 569-1000
Facsimile: (312) 569-3000
carrie.beyer@faegredrinker.com

Elizabeth M.C. Scheibel (admitted pro hac vice)
2200 Wells Fargo Center, 90 S. 7th Street
Minneapolis, MN 55402
Telephone: (612) 766-7000
Facsimile: (612) 766-1600
elizabeth.scheibel@faegredrinker.com

Andrew Lawrence Van Houter
600 Campus Drive
Florham Park, NJ 07932
Telephone: (973) 549-7018
Facsimile: (973) 360-9831
Andrew.vanhouter@faegredrinker.com

Attorneys for Defendant Microsoft Corporation

CERTIFICATE OF WORD COUNT COMPLIANCE

Pursuant to Local Civil Rule 7.1(c), the foregoing Memorandum of Law complies with the Court's order of August 7, 2026 (ECF 1664), as it contains 8,693 words, according to the word processing program used to prepare it.

/s/ Annette L. Hurst
Annette L. Hurst
Attorney for Defendant
Microsoft Corporation