UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OpenAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: 1:23-cv-11195-SHS-OTW 1:24-cv-01515-SHS-OTW 1:24-cv-03285-SHS-OTW 1:24-cv-04872-SHS-OTW 1:25-cv-04315-SHS-OTW	25-md-3143 (SHS) (OTW)

DEFENDANTS' RESPONSE TO NEWS PLAINTIFFS' RULE 56.1 STATEMENT

VOLUME II – Facts 548-1174

FILED UNDER SEAL

VOLUME II

TABLE OF CONTENTS

II. Defendants' Generative AI Products 380

A. Overview of At-Issue LLMs and Products 383

1. Defendants' Large Language Models 383

2. OpenAI's Products 387

3. Microsoft's Products 396

B. Commercial Development and Revenue for Defendants' Products 411

1. OpenAI's Commercial Development and Revenue 411

2. Microsoft's Commercial Development and Revenue 426

C. OpenAI and Microsoft's Relationship and Collaboration 436

1. Collaboration in Model Training 436

2. Financial Relationship Between Microsoft and OpenAI 493

3. Microsoft's Ability to Stop OpenAI's Infringement 505

III. Defendants' Use and Copying of News Content 513

A. Acquisition of News Content 514

1. WebText and WebText2 526

2. Common Crawl 532

3. Project Mango Crawl 537

4. Distribution – Taxi and transfer of GPT-3 training data 538

B. Use of News Content for Data Processing 539

C. Use of News Content During Training 548

1. Use of News Content During Pre-Training 548

2. Use of News Content During Mid-Training 560

3. Use of News Content for Post-Training 570

4. "Memorization" of News Content During Training 586

D. Use of News Content in RAG.....616

1. OpenAI's Use of News Content for RAG.....618

2. Microsoft's Use of News Content for RAG.....644

E. Use of News Content in Outputs.....648

1. OpenAI's Use of News Content in Outputs 653

2. Microsoft's Use of News Content in Outputs 692

F. Defendants' Other Uses of News Content: New York Times Annotated Corpus, Anti-Regurgitation, Project Giraffe 696

1. OpenAI's Use of the New York Times Annotated Corpus to Develop a Commercial Product, in Violation of the User License Agreement 696

2. OpenAI's Use of News Content for Project Giraffe 718

Pursuant to Local Civil Rule 56.1, Defendants OpenAI and Microsoft respectfully submit this Response to Plaintiffs' Corrected Statement of Undisputed Material Facts in Support of their Motion for Partial Summary Judgment. ECF No. 1985 (News) (the "Statement").

Plaintiffs' Statement fails to comply with Local Civil Rule 56.1(a). That rule requires that a motion for summary judgment "be accompanied by a separate, **short, and concise statement**, in numbered paragraphs, of the **material facts** as to which the moving party contends there is no genuine issue to be tried." Local Civ. R. 56.1(a) (emphasis added). That is not what Plaintiffs did. Plaintiffs submitted 940 [*1,825*] separately numbered paragraphs spanning nearly 160 [*440*] pages. Such a voluminous submission does not comply with the local rule. *Tripathy v. McCloskey*, No. 21-CV-6584 (CS), 2024 WL 2135623, at *1 (S.D.N.Y. May 13, 2024) (striking plaintiff's 356-page submission for "ignor[ing] the Rule's requirement that a statement be short and concise"). Plaintiffs' approach has made it extremely difficult for Defendants to respond fully while also providing their responses in a format useful to the Court.

To deal with this massive and improper Statement, Defendants have attempted to streamline their response for the benefit of the Court as follows:

Undisputed/Disputed/Disputed in Part. Every fact includes one of these responses, to make clear Defendants' position under Local Rule 56.1(b). Facts designated "Undisputed" are facts that Defendants have elected not to dispute solely for purposes of Plaintiffs' Motion for Summary Judgment in this matter, but does not constitute a binding admission for any other purpose. Facts designated "Disputed" or "Disputed in Part" are facts that Defendants are disputing.

In addition, Defendants have included certain objections and other responses to Plaintiffs' facts where appropriate. Those additional responses and the authority supporting those responses are identified below. For the Court's convenience and to avoid further lengthening this response, Defendants do not repeat their supporting authority for these issues in every response. Instead, Defendants state these objections plainly for ease of review and incorporate the below authority into each of those responses. Where warranted and helpful for the Court, Defendants also provide additional, paragraph-specific information to assist the Court's assessment.

Immaterial. Facts designated "Immaterial" are facts that, even if true, do not help resolve any question currently before the Court on summary judgment, i.e., they do not "affect the outcome of the suit under the governing law." *I.M. v. United States*, 362 F. Supp. 3d 161, 190 (S.D.N.Y. 2019). Rule 56.1 statements are "immaterial" if they "do not provide any basis to rule in favor of either party for summary judgment even if they are disputed." *See Parks v. Lebhar-Friedman, Inc.*, No. 04-cv-7133-DCP-KNF, 2008 WL 3833802, at *4 (S.D.N.Y. Aug. 11, 2008). Facts designated as "Immaterial" below thus reflect facts that are not legally significant because it does not impact an element of any claim or defense at issue in Plaintiffs' summary judgment motion. Defendants' designation of certain facts as "immaterial" does not concede that any fact not designated as such is material for purposes of adjudicating Plaintiffs' Motion, or for any other purpose.

"Not Supported by Cited Evidence." Facts designated as "Not Supported by Cited Evidence" are facts with incorrect citations or where the cited evidence does not support the stated fact. *Holtz*, 258 F.3d at 74 (A "Rule 56.1 statement is not itself a vehicle for making factual assertions that are otherwise unsupported in the record. Where ... the record does

not support the assertions in a Local Rule 56.1 statement, those assertions should be disregarded and the record reviewed independently.").

"Not Supported by Admissible Evidence." Facts designated as "Not Supported by Admissible Evidence" are facts that are not supported by admissible evidence as required by Local Rule 56.1, and therefore are not properly before the Court. S.D.N.Y. Local Rule 56.1 (d) ("Each statement by the movant … must be followed by citation to evidence that would be admissible[.]"); *Epstein v. Kemper Ins. Companies*, 210 F. Supp. 2d 308, 314 (S.D.N.Y. 2002) (statements in a Rule 56.1 statement "are inappropriate if they are not based on personal knowledge, contain inadmissible hearsay, … or do not cite to supporting evidence").

"Improper Legal Argument." Facts designated as an "Improper Legal Argument" are facts where the assertion therein constitutes improper attorney argument and/or is conclusory. These facts require no response. *Id.* (statements in a Rule 56.1 statement "are inappropriate if they are conclusory or argumentative[.]"); *In re LIBOR-Based Fin. Instruments Antitrust Litig.*, 801 F. Supp. 3d 330, 369 (S.D.N.Y. 2025) (quoting *Kesner v. Buhl*, 590 F. Supp. 3d 680, 691 (S.D.N.Y. 2022)) (the Court sets aside "[t]he portions of [a responding party's] 56.1 Statement that contain legal argument bereft of factual matter[.]").

"Incomplete." Facts designated as "Incomplete" are facts where Plaintiffs have "introduce[d] all or part of a statement" and Defendants introduce "any other part—or any other statement—that in fairness ought to be considered at the same time" pursuant to Federal Rule of Evidence 106.

II. Defendants' Generative AI Products

A. Overview of At-Issue LLMs and Products

1. Defendants' Large Language Models

548. OpenAI released the GPT-1 large language model in June 2018. Ex. 482 [Improving Language Understanding by Generative Pre-Training, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf]; Ex. 483 [Callison-Burch Opening Rep.] ¶ 177.	548. **Undisputed.** Microsoft defers to OpenAI.
549. OpenAI announced the GPT-2 large language model on February 14, 2019. Ex. 484. [Feb. 14, 2019 OpenAI Blog Post, "Better Language Models and Their Implications," https://openai.com/index/better-language-models/] at 1.	549. **Undisputed.** Microsoft defers to OpenAI.
550. OpenAI released GPT-2 in two stages, with an initial release on February 14, 2019, and a full release of its 1.5 billion parameter model on November 5, 2019. Ex. 485 [November 5, 2019 OpenAI Blog Post, "GPT-2: 1.5B release," https://openai.com/index/gpt-2-1-5b-release]; Ex. 484, [Feb. 14, 2019 OpenAI Blog Post, "Better Language Models and Their Implications," https://openai.com/index/better-language-models/].	550. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 550 are disputed because they are not supported by the cited evidence. Pursuant to Fed. R. Evid. 106, there were more than 2 stages to the GPT-2 release (124M in February 2019, 355M in May 2019, 774M in August 2019, and then 1.5B in November 2019). *See* A. Blanco Opp. Decl. Ex. 95 (GPT-2: 6-month follow up, https://openai.com/index/gpt-2-6-month-follow-up); News Ex. 485 (GPT-2: 1.5B release, https://openai.com/index/gpt-2-1-5b-release/).

	Microsoft defers to OpenAI.
551. OpenAI announced the GPT-3 large language model on May 28, 2020, in a paper titled "Language Models are Few-Shot Learners." Ex. 486 [Tom B. Brown, et al., Language Models Are Few-Shot Learners, OpenAI, May 28, 2020, https://arxiv.org/pdf/2005.14165].	551. **Undisputed.** Microsoft defers to OpenAI.
552. OpenAI released GPT-3 on June 11, 2020, and provided access through an API. Ex. 464 [June 11, 2020 OpenAI Blog Post, "OpenAI API," https://openai.com/index/openai-api/.]	552. **Undisputed.** Microsoft defers to OpenAI.
553. ChatGPT, OpenAI's public-facing Chatbot, was released on November 30, 2022. Ex. 465 [November 30, 2022 OpenAI blog post, "Introducing ChatGPT," https://openai.com/index/chatgpt/] at 1; Ex. 466 [Schulman Depo. Tr.] at 162:5-12.	553. **Undisputed.** Microsoft defers to OpenAI.
554. The GPT-3.5 model powered the initial version of ChatGPT. Ex. 465 [November 30, 2022 OpenAI blog post, "Introducing ChatGPT," https://openai.com/index/chatgpt/] at 3; Ex. 466 [Schulman Depo. Tr.] at 162:5-12; Ex. 467 [Goldfarb NYT Opening Rep.] ¶ 62; Ex. 474 [Goldfarb ZD Opening Rep.] ¶ 57.	554. **Undisputed.** Microsoft defers to OpenAI.
555. OpenAI released the GPT-4 large language model on March 14, 2023. Ex. 460 [March 14, 2023 OpenAI Blog Post "GPT-4", https://openai.com/index/gpt-4-research/]; Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 49.	555. **Undisputed.** Microsoft defers to OpenAI.

556. OpenAI released the GPT-3.5 Turbo large language model in March 2023. Ex. 462 [OpenAI's Supplemental Response to NYT Interrogatory No. 7]; Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 50.	556. **Undisputed.** Microsoft defers to OpenAI.
557. Microsoft's Copilot product used a version of GPT-3.5-Turbo called Deucalion from February 2024 through October 2024. Ex. 510 [Microsoft's Supplemental Responses to Daily News Plaintiffs' Interrogatory No. 3].	557. **Undisputed.**
558. Microsoft fine-tuned the Deucalion model. Ex. 510 [Microsoft's Supplemental Responses to Daily News Plaintiffs' Interrogatory No. 3].	558. **Undisputed.**
559. OpenAI released the GPT-4 Turbo large language model in November 2023. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 50.	559. **Undisputed.** Microsoft defers to OpenAI.
560. OpenAI released the GPT-4o large language model in May 2024. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 50.	560. **Undisputed.** Microsoft defers to OpenAI.
561. OpenAI released the GPT-4o mini large language model in July 2024. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 50.	561. **Undisputed.** Microsoft defers to OpenAI.
562. GPT-3 became available in the OpenAI API in July 2020; GPT-3.5 became available in the OpenAI API in March 2022; GPT-4 became available in the OpenAI API in in March 2023; GPT-3.5 Turbo became available in the OpenAI API in March 2023; GPT-4 Turbo became available in the OpenAI API in November 2023; GPT-4o became	562. **Undisputed.** Microsoft defers to OpenAI.

available in the OpenAI API in May 2024. Ex. 462 [OpenAI's Supplemental Response to NYT Interrogatory No. 7].	
563. OpenAI trained GPT-1, GPT-2, GPT-3, GPT-3.5, GPT-4, GPT-3.5 Turbo, GPT-4 Turbo, GPT-4o, and GPT-4o mini. Ex. 482 [Improving Language Understanding by Generative Pre-Training, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf]; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 177, 180, 188-89, 254-55, 257, 258, 260, 263; Ex. 484 [Feb. 14, 2019 OpenAI Blog Post "Better Language Models and Their Implications", https://openai.com/index/better-language-models/]; Ex. 486 [Tom B. Brown et al., Language Models Are Few-Shot Learners, OpenAI (May 28, 2020, https://arxiv.org/pdf/2005.14165)]; Ex. 487 [May 28, 2020 OpenAI Blog Post "Language Models Are Few-Shot Learners", https://openai.com/index/language-models-are-few-shot-learners/]; Ex. 460 [March 14, 2023 OpenAI Blog Post "GPT-4", https://openai.com/index/gpt-4-research/]; Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶¶ 49-50, 64; Ex. 729 [Aug. 22, 2023 OpenAI Blog Post "GPT-3.5 Turbo Fine-Tuning and API Updates", https://openai.com/index/gpt-3-5-turbo-fine-tuning-and-api-updates/].	563. **Undisputed**. Microsoft defers to OpenAI.
564. GPT-2, GPT-3, GPT-3.5, GPT-4, GPT-3.5 Turbo, GPT-4 Turbo, GPT-4o, and GPT-4o mini were trained on News Plaintiffs' content. Ex. 488 [OpenAI's June 10, 2025	564. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI's discovery responses speak for themselves and state that OpenAI located certain

Supplemental Response to NYT Interrogatory No. 7]; Ex. 489 [OpenAI's January 27, 2026 Supplemental Response to ZD Interrogatory No. 6] at 16; Ex. 490 [OpenAI's Second Supplemental Response to Intercept Interrogatory No. 7 (Mar. 23, 2026)] at 6-7; Ex. 491 [OpenAI's Second Supplemental Response to CIR Interrogatory No. 14 (Mar. 23, 2026)] at 6-7; *See* Delorey Opening Report ¶¶ 30, 173, 218–219, 289-90; Ex. 492 [OpenAI's February 4, 2026 Supplemental Responses to News Plaintiffs' Request for Admission No. 69]; Goldstein Opening Rep. ¶ 261; Ex. 9 [Sohl Dep. Ex. 410]; Ex. 493 [OPCO_MDL_000133274] at 74–76; *see also* Facts 1186-1187, 1212, 1222 *infra*.	Plaintiffs' web domains in the contents of certain relevant training datasets affiliated with the stated models. *See* News Ex. 488 (OpenAI's June 10, 2025 Supp. Resp. to NYT Interrog. No. 7) at 8; News Ex. 489 (OpenAI's Jan. 27, 2026 Supp. Resp. to ZD Interrog. No. 6) at 15-16; News Ex. 490 (OpenAI's Mar. 23, 2026 Second Supp. Resp. to Intercept Interrog. No. 7) at 7; News Ex. 491 (OpenAI's Mar. 23, 2026 Second Supp. Resp. to CIR Interrog. No. 14) at 7. However, disputed to the extent that Plaintiffs' characterization of, and inferences drawn from, the cited evidence suggest that every News Plaintiff's content was located in the training datasets for all stated models which is not supported by the evidence. For those reasons, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.

2. OpenAI's Products

i) ChatGPT

565. ChatGPT is a text-generating chatbot that provides natural language responses to user prompts. Ex. 465 [November 30, 2022 OpenAI blog post, "Introducing ChatGPT," https://openai.com/index/chatgpt/] at 1.	565. **Undisputed.** Microsoft defers to OpenAI.
566. OpenAI developed ChatGPT and released it to the public in November 2022. Ex. 465 [November 30, 2022 OpenAI blog post, "Introducing ChatGPT,"	566. **Undisputed.** Microsoft defers to OpenAI.

https://openai.com/index/chatgpt/] at 1; Ex. 466 [Schulman Depo. Tr.] at 162:5-12.	
567. When ChatGPT was first released in November 2022, it was powered by the GPT-3.5 large language model. Ex. 465 [November 30, 2022 OpenAI blog post, "Introducing ChatGPT," https://openai.com/index/chatgpt/] at 3; Ex. 466 [Schulman Depo. Tr.] at 162:5-12; Ex. 467 [Goldfarb NYT Opening Rep.] ¶ 62; Ex. 474 [Goldfarb ZD Opening Rep.] ¶ 57.	567. **Undisputed.** Microsoft defers to OpenAI.
568. Over time, ChatGPT has been powered by different GPT large language models, including GPT-3.5 Turbo, GPT-4 and GPT-4o. Ex. 494 [ChatGPT Release Notes, https://help.openai.com/en/articles/6825453-chatgpt-release-notes]; Ex. 483 [Callison-Burch Opening Rep. ¶¶ 254–64].	568. **Undisputed.** Microsoft defers to OpenAI.
569. In May 2023, OpenAI released the "Browse with Bing" feature of ChatGPT, which was described as a "new version of ChatGPT that knows when and how to browse the internet to answer questions about recent topics and events." Ex. 494 [ChatGPT Release Notes, https://help.openai.com/en/articles/6825453-chatgpt-release-notes].	569. **Undisputed.** Microsoft defers to OpenAI.
570. In October 2024, OpenAI released the "ChatGPT Search" feature of ChatGPT, describing it as a feature that provides "fast, timely answers with links to relevant web sources, which you would have previously needed to go to a	570. **Undisputed.** Microsoft defers to OpenAI.

search engine for." Ex. 495 [Oct. 31, 2024 OpenAI Blog Post, "Introducing ChatGPT search," https://openai.com/index/introducing-chatgpt-search/.]	
571. This "ChatGPT Search" feature replaced the "Browse with Bing" feature of ChatGPT. Ex. 136 [Shetty 30(b)(6) Depo. Tr.] at 99:17-22.	571. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed to the extent Plaintiffs' statement suggests that Browse and Search are two different features. The disputed proposition is not supported by the cited evidence. Specifically, the cited deposition testimony shows that Search is an updated version of Browse: "Browse was an initial version of this product, which was later replaced by search." *See* A. Blanco Opp. Decl. Ex. 276 (V. Shetty 30(b)(6) Depo. Tr.) 99:20-22. Microsoft defers to OpenAI.
572. This "ChatGPT Search" feature of ChatGPT remains active today. Ex. 409 [Perelman 30(b)(6) Depo. Tr.] at 24:14-15.	572. **Undisputed.** Microsoft defers to OpenAI.
573. OpenAI has also released a ChatGPT Commerce feature, which offers users the opportunity to purchase products from retailers within the ChatGPT Interface. Ex. 496 [Dyett Depo. Tr.] at 81:16-23.	573. **Undisputed.** Microsoft defers to OpenAI.
574. OpenAI released this ChatGPT Commerce feature in 2024. Ex. 497 [OPCO_MDL_001697051] at -51.	574. **Disputed.** OpenAI responds as follows: Disputed that the cited evidence establishes that OpenAI released "ChatGPT Commerce" in 2024. News Ex. 497 does not state that ChatGPT Commerce was released in 2024; it reflects plans concerning the launch of Browse in

	2024. Further disputed because Plaintiffs' own Ex. 494 reflects that the ChatGPT Commerce feature was released in Fall 2025, not 2024. News Ex. 494 (ChatGPT – Release Notes) at 65. Microsoft defers to OpenAI.
575. By February 2026 there were 900 million weekly ChatGPT users. Ex. 498 [Goldfarb Reply Rep. ¶ 169]; Ex. 499 [Goldfarb ZD Reply Rep.] ¶ 135.	575. **Undisputed.** Microsoft defers to OpenAI.
576. There are paid versions of ChatGPT, including ChatGPT Plus and ChatGPT Pro, for which OpenAI charges subscription fees. Ex. 500 [May 23, 2025 Monaco Declaration] ¶ 2.	576. **Undisputed.** Microsoft defers to OpenAI.
577. OpenAI generates revenue from subscription fees for paid versions of ChatGPT. Ex. 501 [Lightcap 30(b)(6) Depo. Ex. 1].	577. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI generates revenue from subscription fees for paid versions of ChatGPT. However, disputed to the extent Paragraph 577 implies that OpenAI has always generated revenue from subscription fees for paid versions of ChatGPT. News Ex. 501 reflects all revenue for ChatGPT products beginning in 2023, not the entire period implied by Paragraph 577. *See* News Ex. 501 at 1. Microsoft defers to OpenAI.
578. There is an enterprise version of ChatGPT, for which OpenAI charges subscription fees. Ex. 502 [August 28, 2023 OpenAI Blog Post, "Introducing ChatGPT	578. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that there is an enterprise version of ChatGPT for which OpenAI

Enterprise," https://openai.com/index/introducing-chatgpt-enterprise/]; Ex. 501 [Lightcap 30(b)(6) Depo. Ex. 1].	charges subscription fees. However, disputed to the extent Paragraph 578 implies that OpenAI has always charged subscription fees for ChatGPT Enterprise. Disputed that the cited evidence supports an unqualified statement as to the full relevant period, because News Ex. 502 shows that ChatGPT Enterprise was launched in August 2023, and News Ex. 501 reflects ChatGPT Enterprise revenue only beginning in 2023. *See* News Ex. 502 (Aug. 28, 2023 OpenAI Blog Post, "Introducing ChatGPT Enterprise," https://openai.com/index/introducing-chatgpt-enterprise/); Ex. 501 (Lightcap 30(b)(6) Depo. Ex. 1) Microsoft defers to OpenAI.
579. OpenAI generates revenue from subscription fees from the enterprise version of ChatGPT. Ex. 502 [August 28, 2023 OpenAI Blog Post, "Introducing ChatGPT Enterprise," https://openai.com/index/introducing-chatgpt-enterprise/]; Ex. 501 [Lightcap 30(b)(6) Depo. Ex. 1].	579. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI generates revenue from subscription fees from the enterprise version of ChatGPT. However, disputed to the extent Paragraph 579 implies that OpenAI has always generated revenue from subscription fees from the enterprise version of ChatGPT. Disputed that the cited evidence supports an unqualified statement as to the full relevant period, because News Ex. 502 shows that ChatGPT Enterprise was launched in August 2023, and News Ex. 501 reflects ChatGPT Enterprise revenue only beginning in 2023. *See* Ex. 502 (Aug. 28, 2023 OpenAI Blog Post, "Introducing ChatGPT Enterprise," https://openai.com/index/introducing-chatgpt-enterprise/); News Ex. 501 (Lightcap 30(b)(6) Depo. Ex. 1).

	Microsoft defers to OpenAI.
580. As of February 27, 2026, OpenAI had at least 50 million paid users for consumer ChatGPT accounts, and 9 million paying business users for work and government ChatGPT accounts. Ex. 503 [February 27, 2026 OpenAI Blog Post, "Scaling AI for Everyone," https://openai.com/index/scaling-ai-for-everyone/] at 1.	580. **Disputed in Part – Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI had "more than 50 million consumer subscribers." News Ex. 503 (OpenAI Blog Post, "Scaling AI for Everyone" (Aug. 19, 2026)) at 2. However, disputed that the figures come from February 27, 2026. The blog post contains an August 19, 2026 date. *Id.*Further disputed that there are "9 million paying business users for work and government ChatGPT accounts." News Ex. 503 states: "More than 9 million paying business users rely on ChatGPT for work, and startups, enterprises, and government are building on the OpenAI platform to transform how their products and services are designed, delivered, and run." News Ex. 503 at 1. Microsoft defers to OpenAI.
581. In 2024 OpenAI generated over [redacted] in ChatGPT Plus subscription fees. Ex. 504 [OPCO_MDL_004754941] at "Annual Revenue by Product" Tab.	581. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows. Undisputed that News Ex. 504 shows that OpenAI generated over [redacted] from ChatGPT Plus. However, disputed because News Ex. 504 does not specify that such revenue is from subscription fees. Microsoft defers to OpenAI.

ii) The OpenAI API

582. OpenAI also developed the OpenAI API, which was commercially released to the public in June 2020. Ex. 464 [June 11, 2020 OpenAI Blog Post, "OpenAI API," https://openai.com/index/openai-api/] at 1; Ex. 505 [Lightcap 30(b)(6) Depo. Tr.] at 15:5–10.	582. **Undisputed.** Microsoft defers to OpenAI.
583. OpenAI's former Chief Technology Officer Mira Murati testified at deposition that the OpenAI API is a "commercial" product. Ex. 506 [Murati Depo. Tr.] at 116:3–5.	583. **Disputed in Part – Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that Plaintiffs' characterization that "Mira Murati testified at deposition that the OpenAI API is a 'commercial' product" is supported by the cited evidence. Pursuant to Fed. R. Evid. 106, Ms. Murati actually testified that [redacted] Hurst Decl. Ex. 80 (M. Murati 30(b)(1) Tr.) 116:3–5. Microsoft defers to OpenAI.
584. The OpenAI API enables developers to access OpenAI's LLMs, including to integrate the API into their own products. Ex. 464 [June 11, 2020 OpenAI Blog Post, "OpenAI API," https://openai.com/index/openai-api/] at 1.	584. **Undisputed.** Microsoft defers to OpenAI.
585. At the time it was released, the OpenAI API enabled users to access the GPT-3 large language model. Ex. 464 [June 11, 2020 OpenAI Blog Post, "OpenAI API," https://openai.com/index/openai-api/] at 2; Ex. 505 [Lightcap 30(b)(6) Depo. Tr.] at 15:5–10; Ex. 488 [OpenAI's June 10, 2025 Supplemental Response to NYT Interrogatory No. 7] at 6–7.	585. **Undisputed.** Microsoft defers to OpenAI.

586. Since then, subsequently released GPT models have been made available in the OpenAI API, including GPT 3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, and GPT-4o. Ex. 488 [OpenAI's June 10, 2025 Supplemental Response to NYT Interrogatory No. 7] at 6–7.	586. **Undisputed.** Microsoft defers to OpenAI.
587. OpenAI earns revenue from the OpenAI API, including by charging developers who access and use OpenAI's large language models through the OpenAI API. Ex. 505 [Lightcap 30(b)(6) Depo. Tr.] at 14:6–10.	587. **Undisputed.** Microsoft defers to OpenAI.

iii) CustomGPTs or "GPTs"

588. In November 2023, OpenAI announced CustomGPTs, i.e., "custom versions of ChatGPT that [users] can create for a specific purpose." Ex. 507 [Nov. 6, 2023 OpenAI Blog Post "Introducing GPTs," https://openai.com/index/introducing-gpts/] at 1.	588. **Undisputed.** Microsoft defers to OpenAI.
589. CustomGPTs or "GPTs" enable "anyone to create a tailored version of ChatGPT to be more helpful in their daily life, at specific tasks, at work, or at home—and then share that creation with others." Ex. 507 [Nov. 6, 2023 OpenAI Blog Post "Introducing GPTs," https://openai.com/index/introducing-gpts/] at 1.	589. **Undisputed.** Microsoft defers to OpenAI.
590. CustomGPTs are powered by OpenAI's large language models. Ex. 508 [Ramchandani 30(b)(6) Depo. Tr.] at 17:22–18:4.	590. **Undisputed.** Microsoft defers to OpenAI.

591. Users must visit chatgpt.com or use the ChatGPT app to use a CustomGPT. Ex. 508 [Ramchandani 30(b)(6) Depo. Tr.] at 14:22-16:25; Ex. 554 [OpenAI Help Page, "Creating and editing GPTs," https://help.openai.com/en/articles/8554397-creating-and-editing-gpts].	591. **Undisputed.** Microsoft defers to OpenAI.
592. Some CustomGPTs are made available for use by others in OpenAI's GPT Store, which "is a place where users can discover" CustomGPTs. Ex. 508 [Ramchandani 30(b)(6) Depo. Tr.] at 31:21–32:4.	592. **Undisputed.** Microsoft defers to OpenAI.
593. OpenAI's usage policies apply to building and using CustomGPTs. Ex. 508 [Ramchandani 30(b)(6) Depo. Tr.] at 18:10-19:2.	593. **Undisputed.** Microsoft defers to OpenAI.
594. Before a CustomGPT can be shared with other users, OpenAI runs "automated checks" to assess whether the CustomGPT complies with OpenAI's usage policies, and the CustomGPT cannot be shared with others until it passes that assessment. Ex. 508 [Ramchandani 30(b)(6) Depo. Tr.] at 30:4–25.	594. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI runs automated checks before a CustomGPT can be shared with other users. However, dispute that Plaintiffs' characterization that those checks assess whether a CustomGPT "complies with OpenAI's usage policies" accurately reflects Mr. Ramchandani's testimony to the extent it suggests a comprehensive compliance review. Pursuant to Fed. R. Evid. 106, Mr. Ramchandani testified that the automated systems "look at a myriad of pieces from across all parts of our usage policy," but expressly acknowledged that they are "not looking at the entirety of all copyright issues in automated systems." A. Blanco Opp. Decl. Ex. 262 (R. Ramchandani 30(b)(6) Tr.) 77:23–78:14.

	Microsoft defers to OpenAI.
595. OpenAI also permits users to include CustomGPTs in OpenAI's CustomGPT Store, which is a place for OpenAI users to search for and find CustomGPTs they wish to use. Ex. 509 [January 10, 2024 OpenAI Blog Post, "Introducing the GPT Store," https://openai.com/index/introducing-the-gpt-store/].	595. **Undisputed.** Microsoft defers to OpenAI.
596. Before a CustomGPT can be made available for use by others in OpenAI's GPT Store, OpenAI runs additional assessments to assess whether the CustomGPT complies with OpenAI's usage policies, and the CustomGPT cannot be added to the GPT Store until it passes that assessment. Ex. 508 [Ramchandani 30(b)(6) Depo. Tr.] at 31:4-32:23.	596. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI runs automated checks before a CustomGPT can be shared with other users. However, dispute that Plaintiffs' characterization that those checks assess whether a CustomGPT "complies with OpenAI's usage policies" accurately reflects Mr. Ramchandani's testimony to the extent it suggests a comprehensive compliance review. Pursuant to Fed. R. Evid. 106, Mr. Ramchandani testified that the automated systems "look at a myriad of pieces from across all parts of our usage policy," but expressly acknowledged that they are "not looking at the entirety of all copyright issues in automated systems." A. Blanco Opp. Decl. Ex. 262 (R. Ramchandani 30(b)(6) Tr.) 77:23–78:14. Microsoft defers to OpenAI.

3. Microsoft's Products

i) Bing Chat / Copilot

597. Like ChatGPT, Bing Chat is a text-generating chatbot that provides natural language responses to user prompts. Ex. 82 [Tucker Opening Rep.] ¶ 8.	597. **Disputed in Part.** Defendants do not dispute that Bing Chat is a chatbot with text-generating capabilities that can provide natural language responses to user prompts. However, Defendants dispute that the cited support makes any connection between ChatGPT and Bing Chat. Paragraph 8 instead discusses ChatGPT in (a) and separately Bing Chat in (b). News Ex. 82 (Tucker Opening Rep. (News Pltfs)) ¶ 8.
598. Microsoft developed Bing Chat. Ex. 82 [Tucker Opening Rep.] ¶ 8; Ex. 510 [Microsoft's Second Supplemental Response to Daily News Plaintiff's Second Set of Interrogatories], No. 3; Ex. 511 [Microsoft's June 21, 2024 Response to New York Times Interrogatory No. 8].	598. **Undisputed.**
599. Bing Chat was first released in February 2023, and at that point it was powered by the GPT-4 large language model. Ex. 510 [Microsoft's Second Supplemental Response to Daily News Plaintiff's Second Set of Interrogatories], No. 3; Ex. 511 [Microsoft's June 21, 2024 Response to New York Times Interrogatory No. 8].	599. **Undisputed.**
600. Bing Chat was thereafter powered by additional GPT large language models, including GPT-3.5 Turbo, GPT-4 Turbo, and Deucalion. Ex. 510 Microsoft's Second Supplemental Response to Daily News Plaintiff's Second Set of Interrogatories], No. 3; Ex. 517 [Lafferty Opening Rep.] ¶ 185.	600. **Undisputed.**
601. In November 2023, Microsoft renamed and rebranded the Bing Chat product to Microsoft Copilot. Ex.	601. **Undisputed.**

82 [Tucker Opening Rep.] ¶ 8; Ex. 512 [Nov. 15, 2023 Microsoft Blog Post, "Our Vision to Bring Microsoft Copilot to Everyone, and More," https://blogs.bing.com/search/november-2023/our-vision-to-bring-microsoft-copilot-to-everyone-and-more) at 1; Ex. 513 [Ribas Depo. Tr.] at 19:21–23.	
602. Microsoft incorporates OpenAI's IP, such as the models at issue, into existing products to "make them more desirable and allow them to sort of grow more, ultimately resulting in revenue and profit." Ex. 514 [Tinter 30(b)(6) Depo. Tr.] at 53:12–24.	602. **Disputed in Part – Not Supported by Cited Evidence; Incomplete.** Microsoft does not dispute that it incorporates OpenAI's models into certain Microsoft products or that Mr. Tinter made the quoted statement. But Microsoft responds that the fact stated in Paragraph 602 is disputed because the cited evidence does not support the claim. Paragraph 602 cites to Mr. Tinter's transcript (News Ex. 514) at 53:12-24, but the quoted portion appears at 53:4-6. Pursuant to Fed. R. Evid. 106, Mr. Tinter testified in response to counsel's question specifically regarding Microsoft and OpenAI's commercialization goals: "Q. At the time of this amended and restated joint development and collaboration agreement, what were Microsoft and OpenAI hoping to achieve with respect to commercializing the licensed IP? A. …I'm happy to talk about what Microsoft was attempting to achieve, right? And there was essentially—the theory we had was based on two things: One, development of new products based on that IP for those products to ultimately generate revenue and profit for Microsoft. The second was the incorporation of that product—of that technology into our existing products in

	ways that would make those products more desirable and allow them to sort of grow more, ultimately resulting in revenue and profit." A. Blanco Opp. Decl. Ex. 281 (J. Tinter 30(b)(6) Tr.) 52:15-53:6. OpenAI defers to Microsoft's response.

ii) Second Version of Microsoft Copilot

603. In October 2024, Microsoft released an updated version of the Microsoft Copilot chatbot. Ex. 515 [Suleyman Depo. Tr.] at 26:6–9; Ex. 516 [Suleyman Depo. Ex. 3] (Microsoft October 1, 2024 blog post announcing a "refreshed" version of Copilot); Ex. 517 [Lafferty Opening Rep.] ¶ 185.	603. **Undisputed.**
604. This second version of Copilot is also a text-generating chatbot that provides natural language responses to user prompts. Ex. 515 [Suleyman Depo. Tr.] at 26:6–9; Ex. 516 [Suleyman Depo. Ex. 3] (Microsoft October 1, 2024 blog post announcing a "refreshed" version of Copilot); Ex. 517 [Lafferty Opening Rep.] ¶ 185.	604. **Undisputed.**
605. Microsoft also developed this newer version of Copilot. Ex. 515 [Suleyman Depo. Tr.] at 26:6–9; Ex. 516 [Suleyman Depo. Ex. 3] (Microsoft October 1, 2024 blog post announcing a "refreshed" version of Copilot); Ex. 517 [Lafferty Opening Rep.] ¶ 185.	605. **Undisputed.**

606. This second version of Copilot was, as of February 2026, powered by the GPT-4o model. Ex. 515 [Suleyman Depo. Tr.] at 87:24-88:1.	606. **Undisputed.**
607. There is a paid version of Copilot, called "Copilot Pro"; for which users must pay a subscription fee. Ex. 555 [MSFT_AI_MDL_002758540]; Ex. 560 [Manivannan 30(b)(6) Depo. Tr.] at 17:13–17.	607. **Disputed in Part – Not Supported by Cited Evidence.** Microsoft does not dispute that there was a paid consumer offering of Copilot called Copilot Pro and that Microsoft produced revenue for "Copilot Pro subscriptions" from July 2022 through December 2024. News Ex. 555 (MSFT_AI_MDL_002758540), at worksheet "CoP Pro_US_Rev". Microsoft disputes the remaining facts in Paragraph 607 to the extent it purports to state facts about Copilot Pro or revenue from Copilot Pro after December 2024 because the cited evidence does not support any such facts. News Ex. 555 only covers the period July 2022 through December 2024. News Ex. 555 (MSFT_AI_MDL_002758540). OpenAI defers to Microsoft's response.
608. Microsoft generates revenue from Copilot Pro subscription fees. Ex. 555 [MSFT_AI_MDL_002758540] at 16.	608. **Disputed in Part – Not Supported by Cited Evidence**. Microsoft does not dispute that Microsoft generated revenue from Copilot Pro subscription fees from July 2022 through December 2024. News Ex. 555 (MSFT_AI_MDL_002758540) at worksheet "CoP Pro_US_Rev". Microsoft disputes the remaining facts in Paragraph 608 to the extent it purports to state facts about Copilot Pro or revenue from Copilot Pro after December 2024 because the cited evidence does not support any such representations. News Ex. 555 only covers the period July

	2022 through December 2024. News Ex. 555 (MSFT_AI_MDL_002758540). Microsoft also notes that News Ex. 555 is a native Excel file and it is not clear what Plaintiffs are referring to with the pincite "at 16." OpenAI defers to Microsoft's response.

iii) Microsoft 365 Copilot

609. Microsoft 365 (M365) Copilot is a chatbot product that Microsoft released in November 2023. Ex. 511 [Microsoft's June 21, 2024 Response to New York Times Interrogatory No. 8].	609. **Undisputed.**
610. This chatbot product is available to Microsoft enterprise customers, i.e., people who have a work or a school account through Microsoft. Ex. 511 [Microsoft's June 21, 2024 Response to New York Times Interrogatory No. 8]; Ex. 556 [Amir Opening Rep.] ¶ 23.	610. **Undisputed.**
611. Microsoft generates revenue from charging subscription fees for Microsoft 365 Copilot. Ex. 555 [MSFT_AI_MDL_002758540] at Tab 3 (M365 CoP_US_Rev); Ex. 557 [Microsoft's Subscription Plans, https://www.microsoft.com/en-us/microsoft-365/buy/compare-all-microsoft-365-products].	611. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that it has produced revenue from July 2022 through December 2024 for Microsoft 365 Copilot which was an "an offering for commercial customers to add Copilot functionality to M365 productivity apps." News Ex. 555, at worksheet "M365 CoP_US_Rev". Microsoft disputes the remaining facts in Paragraph 611 to the extent it purports to state facts about Microsoft 365 Copilot after

	December 2024 because the cited evidence does not support any such representations. Additionally, Microsoft disputes the facts in Paragraph 611 to the extent it relies on Ex. 557 as immaterial because these facts do not affect the outcome of the suit under the governing law. News Ex. 557 is a screenshot from a Microsoft website with no indication of when that website was accessed. To the extent that this screenshot was taken after December 2024, the description of Microsoft products and prices does not accurately reflect the Microsoft 365 Copilot product during the relevant time period or for which Microsoft produced revenue information in this case, July 2022 through December 2024. OpenAI defers to Microsoft's response.
612. Microsoft developed the M365 Copilot chatbot. Ex. 511 [Microsoft's June 21, 2024 Response to New York Times Interrogatory No. 8].	612. **Undisputed.**
613. When Microsoft 365 Copilot was first released it was powered by the GPT-4 large language model. Ex. 510 [Microsoft's Second Supplemental Response to Daily News Plaintiff's Second Set of Interrogatories], No. 3.	613. **Undisputed.**
614. Microsoft 365 Copilot has thereafter been powered by additional GPT large language models, including GPT-4 Turbo. Ex. 510 [Microsoft's Second	614. **Undisputed.**

Supplemental Response to Daily News Plaintiff's Second Set of Interrogatories], No. 3.	
615. As of July 2026, Microsoft had at least 30 million paid subscribers for Microsoft 365 Copilot. Ex. 558 [Microsoft Fiscal Year 2026 Fourth Quarter Earnings Conference Call, https://www.microsoft.com/en-us/investor/events/fy-2026/earnings-fy-2026-q4] at 3.	615. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that it was reported in a Microsoft earnings call in July 2026 that Microsoft "now ha[s] over 30 million paid Microsoft 365 Copilot seats." News Ex. 558 at 10. Microsoft disputes that Microsoft had at least 30 million "paid subscribers" because the cited evidence does not support that statement. Microsoft also responds that page 3 of News Ex. 558 does not support the statement in Paragraph 615. Additionally, Microsoft responds that the facts in Paragraph 615 as immaterial because these facts do not affect the outcome of the suit under the governing law. News Ex. 558 is a transcript from a Microsoft earnings call in July 2026. Thus, the facts in Paragraph 615 describe the Microsoft 365 Copilot product after the relevant time period and the time period for which Microsoft produced revenue information in this case, July 2022 through December 2024. OpenAI defers to Microsoft's response.
616. In 2025 Microsoft earned approximately [redacted] in revenue from Microsoft 365 Copilot. Ex. 555 [MSFT_AI_MDL_002758540] at Tab 3.	616. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that it earned approximately [redacted] in revenue from Microsoft 365 Copilot in calendar year July 2024 – December 2024, which is the first half of

	Microsoft's fiscal year 2025. This cited evidence does not provide revenue for "2025." Microsoft also disputes any implication that all of this revenue is reasonably related to the alleged infringement and reserves the right to challenge any damages calculation put forth by Plaintiffs on those grounds should this case proceed to trial. *On Davis v. Gap, Inc.*, 246 F.3d 152, 160 (2d Cir. 2001) ("the term "gross revenue" under the statute means gross revenue reasonably related to the infringement, not unrelated revenues"); News Ex. 127 (Rao Rebuttal Rep.) at 13-14, 74-107. Microsoft also disputes Paragraph 616 as immaterial because these facts do not affect the outcome of the suit under the governing law. Microsoft also disputes Paragraph 616 as immaterial because these facts do not affect the outcome of the suit under the governing law. Paragraph 616 appears to relate entirely to Plaintiffs' damages claims. OpenAI defers to Microsoft's response.

iv) Copilot Search

617. Copilot Search is a chatbot product that Microsoft released in April 2025. Ex. 513 [Ribas Depo. Tr.] at 21:21-22:25; Ex. 559 [April 4, 2025 Microsoft Blog Post, "Introducing Copilot Search in Bing", https://blogs.bing.com/search/April-2025/Introducing-Copilot-Search-in-Bing].	617. **Undisputed – Immaterial.** This fact is undisputed but Microsoft responds that the facts set forth in Paragraph 617 are immaterial because these facts do not affect the outcome of the suit under the governing law. Specifically, the Copilot Search product described in Paragraphs 617-620 is not an accused product at issue in

	this litigation. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW). OpenAI defers to Microsoft's response.
618. Copilot Search is a text-generating chatbot that provides natural language responses to user prompts. Ex. 513 [Ribas Depo. Tr.] at 21:21-22:25; Ex. 559 [April 4, 2025 Microsoft Blog Post, "Introducing Copilot Search in Bing", https://blogs.bing.com/search/April-2025/Introducing-Copilot-Search-in-Bing].	618. **Undisputed – Immaterial.** This fact is undisputed but Microsoft responds that the facts set forth in Paragraph 618 are immaterial because these facts do not affect the outcome of the suit under the governing law. Specifically, the Copilot Search product described in Paragraphs 617-620 is not an accused product at issue in this litigation. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW). OpenAI defers to Microsoft's response.
619. Microsoft developed the Copilot Search product. Ex. 513 [Ribas Depo. Tr.] at 21:21-22:25; Ex. 559 [April 4, 2025 Microsoft Blog Post, "Introducing Copilot Search in Bing", https://blogs.bing.com/search/April-2025/Introducing-Copilot-Search-in-Bing].	619. **Undisputed – Immaterial.** This fact is undisputed but Microsoft responds that the facts set forth in Paragraph 619 are immaterial because these facts do not affect the outcome of the suit under the governing law. Specifically, the Copilot Search product described in Paragraphs 617-620 is not an accused product at issue in this litigation. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW). OpenAI defers to Microsoft's response.
620. Copilot Search as of November 2025 was powered by the GPT-40 mini large language model. Ex. 513 [Ribas Tr.] at 21:21-22:25.	620. **Undisputed – Immaterial.** This fact is undisputed but Microsoft responds that the facts set forth in Paragraph 620 are immaterial because these facts do not affect the outcome of the suit under the governing law. Specifically, the Copilot Search product described in

	Paragraphs 617-620 is not an accused product at issue in this litigation. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW). OpenAI defers to Microsoft's response.

v) Additional Microsoft Implementations of the GPT-4o and GPT-4o Mini models.

621. The GPT-4o and GPT-4o mini models have been used by Microsoft to power over [redacted] Microsoft products. Ex. 561 [November 24, 2025 Declaration of Jacob Thornburgh] ¶ 8.	621. **Disputed in Part – Immaterial.** Microsoft does not dispute that the GPT-4o and GPT-4o mini models have been used in over [redacted] Microsoft products. But Microsoft does dispute the claim in Paragraph 621 that GPT-4o and GPT-4o mini have been "used … to power" such products to the extent that this phrase is intended to overstate the number of unique products that may have used these models. News Ex. 561 states that there were "[redacted] *implementations* of the GPT-4o models," which can include "any new deployments … in new or existing products." News Ex. 561 ¶ 7–8 (emphasis added). Microsoft also disputes any implication from the phrase "used … to power" that the purported products wholly depend on the GPT-4o and GPT-4o mini models for such products' capabilities. Microsoft also responds that any deployment of OpenAI's GPT models in the list of [redacted] products in the Thornburgh declarations is immaterial because these facts do not affect the outcome of the suit under the governing law. The list of [redacted] products in the Thornburgh declarations are not Accused Products in this case. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW).

	OpenAI defers to Microsoft's response.
622. One Microsoft product that has been powered by the GPT-4o or GPT-4o mini models is Recipe Tips, which is a product that returns a list of recipes, ingredients, and tips in response to user queries. Ex. 562 [Microsoft's May 12, 2026 Supplemental Response to NYT Interrogatory 8].	622. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that Recipe Tips has used the GPT-4o model. But Microsoft disputes the claim in Paragraph 622 that Recipe Tips "has been powered by … GPT-4o mini" because this claim is not supported by the cited evidence. News Ex. 562 states that, as of May 12, 2026, "Recipe Tips uses GPT-4o" but does not state that Recipe Tips used the GPT-4o mini model at any point. *See* News Ex. 562 at 9. Microsoft also responds that any deployment of OpenAI's GPT models in Recipe Tips is immaterial because these facts do not affect the outcome of the suit under the governing law. Recipe Tips is not Accused Products in this case. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW). OpenAI defers to Microsoft's response.
623. One Microsoft product that has been powered by the GPT-4o or GPT-4o is Bing Shopping, which uses retrieval augmented generation to search the web in order to identify products in response to user shopping queries. Ex. 562 [Microsoft's May 12, 2026 Supplemental Response to NYT Interrogatory 8].	623. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that Bing Shopping used GPT-4o at one point in time. But Microsoft does dispute the claim in Paragraph 623 that Bing Shopping "has been powered by … GPT-4o mini" because this claim is not supported by the cited evidence. News Ex. 562 states, "Bing Shopping no longer uses GPT-4o" but does not state that Bing Shopping used GPT-4o mini at any point. *See* News Ex. 562 at 9. Microsoft also responds that any deployment of OpenAI's GPT models in Bing Shopping is immaterial

	because these facts do not affect the outcome of the suit under the governing law. Bing Shopping is not Accused Products in this case. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW). OpenAI defers to Microsoft's response.
624. One Microsoft product that has been powered by the GPT-4o or GPT-4o is Summarize Top News from SharePoint in Viva Connections; this product summarizes internal company news by pulling news posts from a user's feed and sending the content to an LLM to generate a summary. Ex. 562 [Microsoft's May 12, 2026 Supplemental Response to NYT Interrogatory 8].	624. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that Summarize Top News from SharePoint in Viva Connections used GPT-4o at one point in time. But Microsoft does dispute the claim in Paragraph 624 that Summarize Top News from SharePoint in Viva Connections "has been powered by … GPT-4o mini" because this claim is not supported by the cited evidence. News Ex. 562 states, "Summarize Top News no longer uses GPT-4o" but does not state that Summarize Top News used GPT-4o mini at any point. *See* News Ex. 562. Microsoft also responds that any deployment of OpenAI's GPT models in Summarize Top News from SharePoint in Viva Connections is immaterial because these facts do not affect the outcome of the suit under the governing law. Summarize Top News from SharePoint in Viva Connections is not Accused Products in this case. *See generally* ECF Nos. 1677, 1620; ECF No. 777 (1:24-cv-03285-SHS-OTW). OpenAI defers to Microsoft's response.
625. When assessing how to integrate ChatGPT's capabilities into its products, Microsoft recognized a	625. **Disputed in Part – Immaterial.** Microsoft does not dispute that News Ex. 563 contains the quoted

number of "challenges." Ex. 563 [MSFT_AI_MDL_000666185] at 185-186 ("[redacted]").	language. But Microsoft responds that the facts set forth in Paragraph 625 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the purported "challenges" and/or [redacted] are not defined and bear no apparent relation to this case. Additionally, any potential "challenges" or [redacted] concerning the potential integration of GPT-3.5 into Microsoft's chatbot product, which is a model that was never actually integrated in Microsoft's Bing Chat product (*see* Paragraphs 599-600), is not a fact that pertains to a claim or defense in this litigation. Microsoft further responds that Paragraph 625 is disputed to the extent that News Plaintiffs represent that this statement pertains to the "Additional Microsoft Implementations of the GPT-4o and GPT-4o Mini models" purportedly reflected in Section II(A)(3)(v). News Ex. 563 discusses GPT-3.5, not GPT-4o or GPT-4o mini. OpenAI defers to Microsoft's response.
626. Microsoft acknowledged the [redacted] Ex. 563 [MSFT_AI_MDL_000666185] at 185-186 ("[redacted]"	626. **Disputed in Part – Not Supported by Admissible Evidence; Immaterial; Improper Legal Argument.** Defendants do not dispute that News Ex. 563 contains the quoted language. But Defendants dispute Paragraph 626 to the extent that News Plaintiffs represent that this statement pertains to the "Additional Microsoft Implementations of the GPT-4o and GPT-4o Mini models" purportedly reflected in this Section II(A)(3)(v). News Ex. 563 discusses GPT-3.5, not GPT-4o or GPT-4o mini.

[redacted]").	Defendants also respond that the facts set forth in Paragraph 626 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, this is an internal document [redacted] These are not facts and are immaterial to the issues before the Court. Moreover, Microsoft's alleged "knowledge" is not an element of any claim or defense at issue in News Plaintiffs' summary judgment motion. Further, the facts set forth in Paragraph 626 are disputed because they are not supported by admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Specifically, Exhibit 563 is an unauthenticated internal Microsoft document, not an OpenAI document, and the quoted language is speculation by an unidentified author about [redacted] To the extent offered for the truth of the matter asserted, it is inadmissible hearsay. Fed. R. Evid. 802; Fed. R. Civ. P. 56(c)(2). Moreover, the document's own custodian, Jordi Ribas, testified he did not believe he had ever seen or received the document and did not believe the document was one he or his team authored. A. Blanco Opp. Decl. Ex. 264 (J. Ribas Tr.) 54:7, 56:27–34. The document is therefore unauthenticated and lacks any foundation establishing it reflects OpenAI's actual training practices. In addition, Defendants dispute that "Microsoft acknowledged" the quoted language in Paragraph 626

	because the cited evidence is unauthenticated, lacks foundation, and lacks any indicia that the statements are attributable to Microsoft. Exhibit 563 is a draft document of unknown authorship that uses [redacted] An individual's internal writing of this sort, completely stripped of context, is not the company's view. Further, to the extent Paragraph 626 relies on the document's characterization that ChatGPT "may have used text that was copyrighted" to establish that OpenAI actually infringed any copyright, that characterization is an improper legal conclusion rather than a fact. The Court should disregard this portion of Paragraph 626 because it constitutes improper attorney argument and/or is conclusory, and it thus requires no response.

B. Commercial Development and Revenue for Defendants' Products

1. OpenAI's Commercial Development and Revenue

627. OpenAI was initially formed as a nonprofit organization in late 2015. Ex. 564 [OpenAI Blog Post, "Our Structure," https://openai.com/our-structure/]; Ex. 565, [OpenAI, Inc. Application for Recognition of Exemption (IRS Form 1023), December 8, 2015]; Ex. 682 ["Documents show OpenAI's long journey from nonprofit to $157B valued company," AP News, https://apnews.com/article/openai-nonprofit-application-tax-exempt-status-ebf38549cdf0529fc5ec727e2815ed12 (citing "Our Structure," OpenAI, https://openai.com/our-structure/; OpenAI, Inc. Application for Recognition of Exemption	627. **Undisputed.** Microsoft defers to OpenAI.

(IRS Form 1023), December 8, 2015;"Documents show OpenAI's long journey from nonprofit to $157B valued company," AP News, https://apnews.com/article/openai-nonprofit-application-tax-exempt-status-ebf38549cdf0529fc5ec727e2815ed12)].	
628. OpenAI in December 2015 represented to the IRS that it did "not plan to play any role in developing commercial products or equipment" and that it "intend[ed] to make its research freely available to the public." Ex. 565 [OpenAI, Inc. Application for Recognition of Exemption (IRS Form 1023), December 8, 2015] at 30.	628. **Undisputed.** Microsoft defers to OpenAI.
629. OpenAI on December 11, 2015 stated that its "goal is to advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return." Ex. 566 [Dec. 11, 2015 OpenAI Blog Post, "Introducing OpenAI," https://openai.com/index/introducing-openai/].	629. **Undisputed.** Microsoft defers to OpenAI.
630. In a 2017 journal entry, OpenAI co-founder and President Greg Brockman wrote that "i was deeply motivated by the gazillions" when he founded OpenAI. Ex. 567 [OPCO_MDL_PD_000007967] at -968.	630. **Disputed in Part.** OpenAI responds as follows: Undisputed that News Ex. 567 contains the quoted text. Disputed to the extent the statement is attributed to Brockman or to OpenAI more generally. News Ex. 567 contains statements from multiple individuals. Further disputed that Brockman's personal journal and Plaintiffs' interpretation of Brockman's personal motivations are relevant to the claims or defenses at issue, and OpenAI objects under Fed. R. Evid. 401 and 403. Further disputed as immaterial for the same reason

	and because this fact does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
631. In a 2017 journal entry, Mr. Brockman wrote that "we've been thinking that maybe we should just flip to a for profit," noting that "it would be nice to be making the billions" and that "we just want something that'll get us the money." Ex. 568 [OPCO_MDL_MUSK_000000879] at 879–80.	631. **Disputed in Part - Not Supported by Admissible Evidence.** OpenAI responds as follows: Undisputed that News Ex. 568 contains the quoted text. Disputed that Plaintiffs' misleading characterization of, and inferences drawn from, News Ex. 568 establish the reasons for OpenAI's organizational restructuring. Other portions of the same contemporaneous journal in News Ex. 568 reflect a broader purpose, including that "our point isn't making the discoveries, it's ensuring that it's a public good." Consistent with that context, Brockman testified that OpenAI came to believe that "a for-profit entity of some form would be really the only way to successfully pursue the mission." A. Blanco Opp. Decl. Ex. 176 (G. Brockman 30(b)(1) Tr.) 450:3–6. The disputed proposition is therefore not supported by the cited evidence. Further disputed that this passage from Brockman's personal journal is relevant to the claims or defenses at issue, and OpenAI objects under Fed. R. Evid. 401 and 403. Further disputed as immaterial because this fact does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.

632. In a 2017 journal entry, Mr. Brockman identified "true dangers" that OpenAI's efforts could unleash, such as "unemploying everyone" and "human extinction," noting an "answer" was "need[ed]" "to get the billions." Ex. 569 [OPCO_MDL_PD_000007912] at -913.	632. **Disputed in Part.** OpenAI responds as follows: Undisputed that News Ex. 569 contains the quoted statement. Disputed that Plaintiffs' characterization accurately reflects the context of News Ex. 569. The entry identifies a range of potential AGI risks under the heading "what are the true dangers?," including "unemploying everyone" and "takeoff risk: dramatically more powerful than humans, human extinction." However, Plaintiffs' juxtaposition of those statements with "need to answer to get the billions" misleadingly suggests that Brockman stated that OpenAI needed an "answer" to those dangers in order "to get the billions." This is inaccurate: the entry actually discusses broader questions concerning OpenAI's strategy, organizational structure, fundraising, and how to articulate a sufficiently compelling path forward to potential funders, immediately following "need to answer to get the billions" with "why should people part with their hard-earned money?" News Ex. 569 at OPCO_MDL_PD_000007913. Further disputed to the extent Plaintiffs characterize the identified "dangers" as consequences Brockman believed would occur, rather than potential risks; the entry itself discusses uncertainty concerning AGI risks and expressly qualifies portions of the discussion with "if these assumptions are true." The disputed proposition is therefore not supported by the cited evidence. Further disputed that this passage from Brockman's personal journal is relevant to the claims or defenses at

	issue, and OpenAI objects under Fed. R. Evid. 401 and 403. Further disputed as immaterial because this fact does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
633. Mr. Altman testified in the *Musk v. Altman* litigation that OpenAI in 2017 began planning for "the possibility of forming a for-profit entity to carry on AI-related activities." Ex. 570 [Altman *Musk* Dep. Tr., Dkt. 1162-6] at 303:15-19; 329:20-24.	633. **Undisputed**. Microsoft defers to OpenAI.
634. Mr. Brockman testified in the *Musk v. Altman* litigation that OpenAI in 2017 "concluded" that it would form a for-profit company. Ex. 571 [Brockman *Musk* Dep. Tr., Dkt. 1162-1] at 229:2-18.	634. **Disputed in Part.** OpenAI responds as follows: Undisputed that News Ex. 571 contains the quoted language. However, disputed that Plaintiffs' characterization that OpenAI "concluded" that it would form a for-profit company in 2017 accurately reflects Ex. 571. The cited testimony instead reflects that certain OpenAI personnel concluded that forming a public benefit corporation or B Corp was the best way to pursue OpenAI's mission, not that OpenAI itself made a final decision in 2017 to form a for-profit company. *See* Ex. 571 (*Musk v. Altman* G. Brockman 30(b)(1) Tr.) at 229:2–18. Microsoft defers to OpenAI.
635. An August 2018 internal OpenAI presentation described the company's plans to pursue "commercial opportunities." Ex. 572 [OPCO_MDL_004420637] at -685.	635. **Disputed in Part.** OpenAI responds as follows: Undisputed that the quoted language appears in News Ex. 572. However, disputed because the cited evidence does not describe OpenAI's plans to pursue

	commercial opportunities but rather states that, "[w]hile [the company] *may* pursue various commercial opportunities in this interim period, *it will not be a core focus of the organization and should not be an expectation*." *See* News Ex. 572 (OPCO_MDL_004420637 at -685) (emphases added). Microsoft defers to OpenAI.
636. OpenAI in late 2018 formed a for-profit entity. Ex. 564 [OpenAI Blog Post, "Our Structure," https://openai.com/our-structure/]; Ex. 539 [Korovesis 30(b)(6) Depo. Tr.] at 23:15-21.	636. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI formed a for-profit entity. Pursuant to Fed. R. Evid. 106, however, disputed that the formation occurred in late 2018. *See* News Ex. 564 (OpenAI Blog Post, OpenAI, https://openai.com/our-structure/); A. Blanco Opp. Decl. Ex. 97 (OpenAI LP, OpenAI, https://openai.com/index/openai-lp/). Microsoft defers to OpenAI.
637. In a January 2019 diary entry, Greg Brockman celebrated that OpenAI "actually got the for-profit [they] were dreaming of." Ex. 573 [OPCO_MDL_PD_000007962] at -963 (describing "earthly rewards" and how "money + fame/respect are the two that I think about the most").	637. **Disputed in Part – Immaterial**. OpenAI responds as follows: Undisputed that News Ex. 573 contains the quoted text. Dispute Plaintiffs' characterization of, and inferences drawn from News Ex. 573. Plaintiffs take the quote out of its context, which included a list of several company goals that included: "preserve liberal values in US & other countries that already have them," "distribute the vast benefits of AGI equally," and "spread liberal values around the world." Further disputed that this passage from Brockman's personal journal is relevant to the claims or defenses at

	issue, and OpenAI objects under Fed. R. Evid. 401 and 403. Further disputed as immaterial because this fact does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
638. Microsoft and OpenAI entered into a partnership agreement in July 2019. Ex. 574 [July 22, 2019 Microsoft Company News, "OpenAI Forms Exclusive Computing Partnership with Microsoft to Build New Azure AI Supercomputing Technologies," https://news.microsoft.com/source/2019/07/22/openai-forms-exclusive-computing-partnership-with-microsoft-to-build-new-azure-ai-supercomputing-technologies/]; Ex. 575 [MSFT_AI_MDL_000000345] (Joint Development and Collaboration Agreement) at 1; Ex. 576 [MSFT_AI_MDL_001550855] at 1; *see infra* Facts 669-780 for more information about the Microsoft-OpenAI partnership.	638. **Undisputed.**
639. OpenAI's former Chief Technology Officer, Mira Murati, confirmed at deposition that "one of the purposes of the Microsoft-OpenAI partnership was to commercialize OpenAI's technology." Ex. 506 [Murati Depo. Tr.] at 21:14-17; *see also* Ex. 533 [Nadella Depo. Tr.] at 132:20–133:3 ("We have a commercial relationship with OpenAI and if OpenAI does well, we do well.").	639. **Disputed in Part - Not Supported by Cited Evidence.** Defendants respond as follows: Undisputed that News Ex. 506 contains the quoted testimony. However, disputed to the extent Plaintiffs suggest Murati was describing the purpose of the 2019 Joint Development and Collaboration Agreement. The cited testimony refers generally to the Microsoft-OpenAI "partnership" not the 2019 JDCA or characterizing the purpose of that agreement.

640. "Commercialization" was a priority for OpenAI's executives in 2019. Ex. 534 [MCCAULEY_000000582] at -586; Ex. 506 [Murati Depo. Tr.] at 143:25-149:7.	640. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that Plaintiffs' characterization that "commercialization" was a priority for OpenAI's executives in 2019 is supported by the cited evidence. News Ex. 534 reflects a performance review identifying "successful commercialization (eg, >$100M of revenue in 2021)" among four goals Altman identified for Murati; it does not establish that commercialization was generally a priority for OpenAI's executives in 2019. Murati's testimony further undermines Plaintiffs' characterization: [redacted] A. Blanco Opp. Decl. Ex. 244 (M. Murati 30(b)(1) Tr.) 143:9-11. Microsoft defers to OpenAI.
641. In a performance review of Mira Murati in the second half of 2019, Sam Altman stated "I have not seen your goals yet, but here are four I care most about: 1) Successful Commercialization (e.g., >$100 million of revenue in 2021); 2) Microsoft remaining happy with us and investing the second billion." Ex. 534 [MCCAULEY000000582] at -586; Ex. 506 [Murati Depo. Tr.] at 143:25-149:7.	641. **Disputed in Part.** OpenAI responds as follows: Undisputed that the quoted language appears in News Ex. 534. Disputed that Plaintiffs' characterization that "commercialization" was a priority for OpenAI's executives in 2019 is supported by the cited evidence. News Ex. 534 reflects a performance review identifying "successful commercialization (eg, >$100M of revenue in 2021)" among four goals Altman identified for Murati; it does not establish that commercialization was generally a priority for OpenAI's executives in 2019. Murati's testimony further undermines Plaintiffs'

	characterization: [redacted] A. Blanco Opp. Decl. Ex. 244 (M. Murati 30(b)(1) Tr.) 143:9-11. Microsoft defers to OpenAI.
642. In the same document, Mira Murati wrote that "In the past six months, I have kept executive team informed of all MSFT and commercialization-related matters on a weekly basis via Slack and formally on a monthly basis via email. I have started office hours for commercialization with all of OpenAI every six weeks and shifting to a weekly update email, given that we have started moving much faster." Ex. 534 [MCCAULEY000000582] at -604; Ex. 506 [Murati Depo. Tr.] at 24:14-27:21.	642. **Disputed in Part.** OpenAI responds as follows: Undisputed that the quoted language appears in News Ex. 534. Disputed that Plaintiffs' characterization that "commercialization" was a priority for OpenAI's executives in 2019 is supported by the cited evidence. News Ex. 534 reflects a performance review identifying "successful commercialization (eg, >$100M of revenue in 2021)" among four goals Altman identified for Murati; it does not establish that commercialization was generally a priority for OpenAI's executives in 2019. Murati's testimony further undermines Plaintiffs' characterization: [redacted] A. Blanco Opp. Decl. Ex. 244 (M. Murati 30(b)(1) Tr.) 143:9-11. Microsoft defers to OpenAI.
643. OpenAI employees in October 2019 understood that OpenAI was not permitted to use data it had collected to develop a commercial product. Ex. 535	643. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that Plaintiffs' characterization of what "OpenAI employees"

[OPCO_MDL_001838416] at -18 ("I don't believe we're allowed to use a lot of our current dataset in a commercial product.").	understood about using data OpenAI had collected to develop a commercial product is supported by the cited evidence. News Ex. 535 reflects one employee's qualified statement, "I don't believe we're allowed to use a lot of our current dataset in a commercial product," not an established understanding among OpenAI employees. The cited exchange also includes Alec Radford's response that "it was just ruled last month in a court of California that scraping publicly available data is legal," further demonstrating that the exchange reflected an ongoing discussion rather than a settled understanding regarding permissible uses of OpenAI's data. *See* News Ex. 535 at OPCO_MDL_001838418. Microsoft defers to OpenAI.
644. As of at least December 4, 2019, OpenAI and Microsoft employees were holding "Commercialization Working Sessions." Ex. 536 [OPCO_MDL_004838040] at -040; Ex. 506 [Murati Depo. Tr.] at 143:25-149:7.	644. **Disputed in Part – Not Supported by the Cited Evidence.** Defendants respond as follows: Undisputed that News Ex. 536 contains the phrase "Commercialization Working Sessions." Disputed that Plaintiffs' characterization that OpenAI and Microsoft employees "were holding" such sessions as of December 4, 2019 is supported by the cited evidence. News Ex. 536 reflects a proposed schedule for "Commercialization Working Sessions," but does not establish that the proposed sessions actually occurred. Murati's testimony likewise does not establish that such meetings occurred; when asked whether she had "regular meetings with Microsoft about commercializing OpenAI's technology"

	around this time, she testified, "I do not remember." A. Blanco Opp. Decl. Ex. 244 (M. Murati 30(b)(1) Tr.) 26:17–27:1. Further, [redacted] *Id.* at 143:9–11.
645. By 2020, OpenAI began generating revenue. *See* Ex. 537 [OPCO_MDL_001005468] at -474.	645. **Undisputed.** Microsoft defers to OpenAI.
646. OpenAI reported in 2020 that its "revenue is primarily derived from fees earned from customers accessing the OpenAI API." Ex. 537 [OPCO_MDL_001005468] at -478.	646. **Undisputed.** Microsoft defers to OpenAI.
647. From 2020 to 2024, OpenAI generated approximately [redacted] in revenue from the models and products at issue in this case, including: ChatGPT and the OpenAI API. Ex. 504 [OPCO_MDL_004754941] ("Annual Revenue by Product Tab"); *see also* Malackowski NYT-DNP Opening Rep. at 153 (Figure 38, summarizing evidence); *see also id.* at 547 (Appendix D, Ex. 1).	647. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI generated at least [redacted] in revenue from 2020 to 2024, including from ChatGPT and the OpenAI API. However, disputed that the [redacted] in revenue was generated exclusively from the models and products at issue in this case. The cited evidence does not establish that all the identified revenue is attributable exclusively to the models and products at issue, as opposed to other products or services offered through ChatGPT and the OpenAI API. Microsoft defers to OpenAI.

648. Under the revenue sharing agreement between OpenAI and Microsoft, Microsoft has contributed approximately [redacted] to OpenAI's revenues from 2023 to 2024. Ex. 538 [MSFT_AI_MDL_002758534 ("RevShare to OpenAI")]; *see also* Malackowski NYT-DNP Opening Report p. 151 and Appendix D, Exs. 2.4 and 5.2.	648. **Undisputed**.
649. Additional sources of revenue have included shopping and advertising. Ex. 539 [Korovesis 30(b)(6) Depo. Tr.] at 170:5–184:1; Ex. 577 [Feb. 9, 2026 OpenAI Blog Post, "Testing ads in ChatGPT," https://openai.com/index/testing-ads-in-chatgpt/]; Ex. 578 ["OpenAI's US ad pilot exceeds $100 million in annualized revenue in six weeks," Reuters, March 26, 2026, https://www.reuters.com/business/media-telecom/openais-us-ad-pilot-exceeds-100-million-annualized-revenue-six-weeks-2026-03-26/)]; Ex. 496 [Dyett Depo. Tr.] at 81:16-23.	649. **Disputed.** OpenAI responds as follows: Disputed that OpenAI had generated revenue from shopping and advertising features of its products as of January 21, 2026. Korovesis testified that OpenAI had not generated revenue from shopping or shopping partners and had not generated revenue from advertising. A. Blanco Opp. Decl. Ex. 224 (J. Korovesis 30(b)(6) Tr.) 171:19–23, 172:4–18, 176:25–177:2, 178:14–16, 181:20–22. Dyett's testimony does not establish otherwise; Dyett testified only that OpenAI's Instant Checkout feature permits users to purchase products from retailers without leaving ChatGPT, not that OpenAI generated revenue from that feature. News Ex. 496 (J. Dyett 30(b)(1) Tr.) 81:16-23. Microsoft defers to OpenAI.
650. At a gross profit level OpenAI is profitable, generating [redacted] respectively. Ex. 579 [OPCO_MDL_004416645] at -677.	650. **Disputed in Part.** OpenAI responds as follows: Undisputed that the cited figures appear in News Ex. 579 and are characterized in the exhibit as "Gross Profit." Disputed that these figures represent gross profit under GAAP principles. *See* A. Blanco Opp. Decl. Ex. 224 (J. Korovesis 30(b)(6) Tr.) 155:18-156:22.

	Microsoft defers to OpenAI.
651. In 2025, OpenAI recapitalized as a public-benefit corporation ("PCB"). Ex. 580 [October 28, 2025 OpenAI Blog Post, "Built to benefit everyone," https://openai.com/index/built-to-benefit-everyone/]; Ex. 581 ["OpenAI may move forward with new business structure, partnership with Microsoft, regulators say," AP News, https://apnews.com/article/openai-chatgpt-nonprofit-microsoft-c661df3242766d6b0ddbab401ad1fd84)]; Ex. 582 ["OpenAI Abandons Planned For-Profit Conversion," The Wall Street Journal, https://www.wsj.com/tech/ai/openai-to-become-public-benefit-corporation-9e7896e0].	651. **Disputed in Part.** OpenAI responds as follows: Undisputed that, in 2025, OpenAI conducted a recapitalization that converted its for-profit entity to a public benefit corporation. Disputed that Plaintiffs' characterization that "OpenAI" recapitalized as a PBC accurately describes the transaction to the extent it suggests that OpenAI as a whole converted into a public benefit corporation. Microsoft defers to OpenAI.
652. In 2026, OpenAI expects to generate $2.5 billion in advertising revenue and $100 billion in annual advertising revenue by 2030. Ex. 583 ["OpenAI projects $2.5 billion in ad revenue this year, $100 billion by 2030, Axios reports," Reuters, April 9, 2026, https://www.reuters.com/business/media-telecom/openai-projects-25-billion-ad-revenue-this-year-100-billion-by-2030-axios-2026-04-09/)].	652. **Disputed.** OpenAI responds as follows: Disputed that the cited evidence establishes that OpenAI expects to generate $2.5 billion in advertising revenue in 2026 or $100 billion annually by 2030. News Ex. 583 is a Reuters article reporting projections attributed to another news organization, Axios, rather than a statement or business record from OpenAI. OpenAI objects to Plaintiffs' reliance on News Ex. 583 for the truth of those projections as inadmissible hearsay and for lack of foundation. Fed. R. Evid. 802, 901; Fed. R. Civ. P. 56(c)(2). Further disputed to the extent Plaintiffs characterize third-party reporting concerning projected future revenue as establishing OpenAI's actual expectations or future financial performance.

	Microsoft defers to OpenAI.
653. OpenAI's valuation has increased to at least $852 billion. Ex. 584 ["OpenAI raises $122 billion to accelerate the next phase of AI," OpenAI, March 31, 2026, https://openai.com/index/accelerating-the-next-phase-ai/].	653. **Disputed in Part.** OpenAI responds as follows: Undisputed that, on March 31, 2026, OpenAI closed a funding round with $122 billion in committed capital at a post money valuation of $825 billion. Disputed that the cited evidence supports any broader conclusion about OpenAI's valuation. Microsoft defers to OpenAI.
654. OpenAI is planning a $1 trillion IPO. Ex. 585 ["Exclusive: OpenAI lays groundwork for juggernaut IPO at up to $1 trillion valuation," Reuters, https://www.reuters.com/business/openai-lays-groundwork-juggernaut-ipo-up-1-trillion-valuation-2025-10-29/)].	654. **Disputed.** OpenAI responds as follows: Disputed that the cited evidence establishes that OpenAI is planning a $1 trillion IPO. News Ex. 585 is a Reuters article reporting information attributed to unidentified sources, not a statement or business record from OpenAI. OpenAI objects to Plaintiffs' reliance on News Ex. 585 for the truth of those assertions as inadmissible hearsay and for lack of foundation. Fed. R. Evid. 802, 901; Fed. R. Civ. P. 56(c)(2). Microsoft does not believe that any part of this alleged fact is directed at Microsoft and on this basis defers to OpenAI.
655. Many OpenAI employees have received equity in OpenAI as part of their compensation and for many of these employees, the value of their equity interest in OpenAI is substantial. Ex. 130 [Brockman Depo. Tr.] at 39:11–13 ("[redacted]"); Ex. 233 [Lightcap	655. **Disputed in Part.** OpenAI responds as follows: Undisputed that multiple OpenAI employees have received equity in OpenAI as part of their compensation. Disputed that Plaintiffs' characterization that "many OpenAI employees" have "substantial" equity interests is supported by the cited evidence. The cited

30(b)(1) Depo. Tr.] at 13:4–9 (Brad Lightcap testifying that [redacted]); Ex. 466 [Schulman Depo. Tr.] at 14:13-14 (John Schulman testifying that [redacted]; Ex. 586 [Sutskever *Musk* Depo. Tr.] at 375:6-17 (Ilya Sutskever testifying that [redacted]"); Ex. 524 [Pachocki Depo. Tr.] at 424:4-10 (Jakub Pachocki testifying that the [redacted]); Ex. 506 [Murati Depo. Tr.] at 19:2–3 (Mira Murati testifying that her [redacted] Ex. 525 [Ryder 30(b)(1) Depo. Tr.] at 72:6-10 ("[redacted]"); Ex. 526 [Paino 30(b)(1) Depo. Tr.] at 12:8-9 (Alex Paino testifying that [redacted]; Ex. 527 [Welinder Nov. 13, 2025 Depo. Tr.] at 12:15-19 (Peter Welinder testifying that [redacted]"); Ex. 528 [Turley 30(b)(1) Depo. Tr.] at 17:11-14 (Nick Turley testifying that [redacted]); Ex. 523 [Passos Depo. Tr.] at 221:2-6 (Alex Passos testifying that [redacted].	testimony concerns the equity holdings of a limited group of specifically identified individuals and does not establish the number or proportion of OpenAI employees who hold equity of comparable value or otherwise establish that these individuals are representative of OpenAI employees generally. Microsoft defers to OpenAI.
656. Many OpenAI employees have generated substantial income from selling their equity interest in OpenAI, or portions thereof. Ex. 130 [Brockman Depo. Tr.] at 39:17-20 (Greg Brockman: [redacted]); Ex. 506 [Murati Depo. Tr.] at 19:6-13 (Mira Murati: "[redacted]	656. **Disputed in Part.** OpenAI responds as follows: Undisputed that multiple OpenAI employees have generated income from selling their equity interest, or portions thereof, in OpenAI. Disputed that Plaintiffs' characterization that "many OpenAI employees" have

[redacted]"); Ex. 524 [Pachocki Depo. Tr.] at 424:4-6 (Jakub Pachoki:[redacted]); Ex. 525 [Ryder 30(b)(1) Depo. Tr.] at 72:19-20 (Nick Ryder: [redacted]); Ex. 527 [Welinder Nov. 13, 2025 Depo. Tr.] at 12:25-13:4 (Peter Welinder: "[redacted]"); Ex. 496 [Dyett Depo. Tr.] at 13:22-24 (James Dyett: "[redacted]"); Ex. 529 [Chess Depo. Tr.] at 141:9-10 (Benjamin Chess: "[redacted]"); Ex. 530 [Heidecke Depo. Tr.] at 15:16-18 (Johannes Heidecke: "[redacted]"); Ex. 528 [Turley 30(b)(1) Depo. Tr.] at 17:18-20 (Nick Turley: "[redacted]").	generated substantial income from such sales is supported by the cited evidence. The cited testimony concerns a limited group of specifically identified individuals and does not establish the number or proportion of OpenAI employees who have sold equity, the amounts realized by OpenAI employees generally, or that the identified individuals are representative of OpenAI employees as a whole. Microsoft defers to OpenAI.

2. Microsoft's Commercial Development and Revenue

657. Microsoft has a market capitalization of over $3 trillion. Ex. 531 [Microsoft Stock Price, MarketWatch, https://www.marketwatch.com/investing/stock/msft].	657. **Disputed in Part – Not Supported by Admissible Evidence; Immaterial.** Microsoft does not dispute that it has a market capitalization of over $3 trillion. But the facts set forth in Paragraph 657 are disputed on the grounds that they are not supported by admissible evidence, as required by Local Civil Rule 56.1, and therefore are not properly before the Court. Specifically, the evidence cited is hearsay and lacks foundation because it is a non-party publication offered for the truth of its contents, unaccompanied by any authenticating testimony or sponsoring expert to establish its reliability and Plaintiffs identify no exception to the hearsay rule that would apply. It therefore may not be considered in support of summary judgment.

658. Microsoft reported to investors that the commercial partnership between OpenAI and Microsoft "has been super beneficial" leading to "one of the most, highest valued private companies today," and "that has led to great success for Microsoft. That's led to great success for OpenAI." Ex. 532 ["Microsoft Fiscal Year 2025 First Quarter Earnings Conference Call" October 30, 2024, https://www.microsoft.com/en-us/investor/events/fy-2025/earnings-fy-2025-q1] at 19.	658. **Undisputed.**
659. Microsoft's CEO Satya Nadella admitted at deposition that, from the very beginning of the partnership, Microsoft's purpose for pursuing a partnership with OpenAI was to pursue a commercial relationship. Ex. 533 [Nadella Depo. Tr.] at 209:24-210:4 ("Q: And from Microsoft's perspective, from the very beginning, you weren't making this investment for nonprofit purposes, you were doing it because it was in the best interest of Microsoft's business ambitions, fair? A. Yeah"); *id.* 210:17-21 ("Q: And from the very beginning of those discussions, you understood that Microsoft would be doing it for its own commercial purposes, correct? A. That is correct.").	659. **Undisputed - Incomplete.** Microsoft does not dispute that News Exhibit 533 contains the quoted deposition testimony. But, pursuant to Fed. R. Evid. 106, Mr. Nadella's full testimony includes how Microsoft's purpose for pursuing a partnership with OpenAI was informed by how "OpenAI came to the realization that in order for them to continue to be a lab that is doing leading work, they needed to fund a lot of compute and therefore they needed to create a for-profit entity and find a commercial partner who was willing to take risks, at which we took." *See* A. Blanco Opp. Decl. Ex. 245 (S. Nadella 30(b)(1) Tr.) 209:24-210:10. Microsoft did pursue a commercial relationship in its partnership with OpenAI. *See, e.g.*, Hurst Decl. Ex. 81 (S. Nadella 30(b)(1) Tr. 250:16-251:24 (describing the partnership as Microsoft's "strategic intent . . . to support AI companies in that time. . . . [O]ur entire goal was to sponsor anyone with ambition to compete at the frontier of AI. And that's what we were pushing ourselves to figure out a commercial construct that

	would allow us to do that. . . . In fact, that's sort of what attracted me to OpenAI is that they had, in some sense, the same ambition that we had, which is that we want to build great AI."). OpenAI defers to Microsoft's response.
660. Microsoft generates revenues relating to the models and products at issue through several revenue streams, including: revenue sharing with OpenAI; OpenAI's use of the Microsoft Azure computing infrastructure for inferencing (e.g., operating the models to generate outputs); offering OpenAI's LLMs to third parties through Azure; incorporating OpenAI products into new and existing Microsoft products; use of OpenAI IP in Microsoft's Search, Advertising, News, and Edge ("SANE") products; and sale of the Bing Search API to OpenAI. *See* Malackowski NYT-DNP Opening Rep. at Section 11.1 (collecting evidence); *see also* Facts 662-668, 759 below; Ex. 587 [MSFT_AI_MDL_002881404]; Ex. 538 [MSFT_AI_MDL_002758534]; Ex. 588 [MSFT_AI_MDL_002846147]; Ex. 589 [MSFT_AI_MDL_002881406]; Ex. 555 [MSFT_AI_MDL_002758540]; Ex. 590 [MSFT_AI_MDL_002758528]; Ex. 591 [MSFT_AI_MDL_002758535]; Ex. 592 [MSFT_AI_MDL_002758533].	660. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that it offered, and generated revenue for, the products listed in Paragraph 660 through December 2024 and produced revenue and costs related to those products in this case. Microsoft disputes the remaining facts in paragraph 660 to the extent it purports to state facts about revenue from models and products not at issue in this case or generated after December 2024, because the cited evidence does not support any such statements. The cited evidence is the financial data produced by Microsoft in this case, which runs through December 2024. Microsoft also disputes any implication that each and all of these revenue streams are reasonably related to the alleged infringement and reserves the right to challenge any damages calculation put forth by Plaintiffs on those grounds should this case proceed to trial. *On Davis,* 246 F.3d at 160 ("the term "gross revenue" under the statute means gross revenue reasonably related to the infringement, not unrelated revenues"); News Ex. 127 (Rao Rebuttal Rep.) ¶¶ 9 (at pp. 13-14), 117-157.

	Microsoft also disputes the facts in paragraph 660 as immaterial because these facts do not affect the outcome of the suit under the governing law. Paragraph 660 appears to relate entirely to Plaintiffs' damages claims. Additionally, Microsoft disputes Paragraph 660 as immaterial to the extent it states facts regarding products that are not an accused product at issue in this litigation. *See generally* MDL ECF Nos. 1677 (NYT Third Am. Compl.), 1620 (CIR Second Am. Compl.); ECF No. 777 (1:24-cv-03285-SHS-OTW) (Daily News Pltfs. First Am. Compl.). OpenAI defers to Microsoft's response.
661. Microsoft receives a 20% share of OpenAI API revenue and ChatGPT revenue. Ex. 593 [MSFT_AI_MDL_000000972] at 5; Ex. 601 [Griffith Depo. Tr.] at 53:11–54:14.	661. **Disputed in Part – Not Supported by Cited Evidence; Not Supported by Admissible Evidence.** Defendants do not dispute that Microsoft received a revenue share for a portion of OpenAI's revenue from January 2023 through December 2024. News Ex. 538 (MSFT_AI_MDL_002758534). Defendants note that the revenue that Microsoft received from OpenAI is for only the OpenAI API and ChatGPT [redacted] News Ex. 538. Additionally, [redacted] A. Blanco Opp. Decl. Ex. 281 (J. Tinter 30(b)(6) Tr.) 210:23-212:12, 214:1-215:24 (discussing News Ex. 538). Microsoft disputes the remaining facts in paragraph 661 to the extent it purports to state facts about agreements

	between Microsoft and OpenAI regarding a revenue share before January 2023 or after December 2024, because the cited evidence does not support any such representations. Microsoft also disputes the facts in Paragraph 661 to the extent it relies on Ex. 593 on the grounds that it is not supported by admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Specifically, the cited evidence is hearsay and lacks foundation. Microsoft's 30(b)(6) witness on financials testified that she did not know who put this presentation together and "ha[d]n't seen this document before[,]" and referred to the financial data produced in the case as the source for financial data on the revenue share. A. Blanco Opp. Decl. Ex. 209 (K. Griffith 30(b)(6) Tr.) 52:20-53:9, 56:22-57:3, 63:5-18, 78:4-82:14. OpenAI defers to Microsoft's response.
662. From 2022 to 2024, Microsoft generated approximately [redacted] in revenue relating to the models and products at issue. Ex. 587 [MSFT_AI_MDL_002881404]; Ex. 538 [MSFT_AI_MDL_002758534]; Ex. 588 [MSFT_AI_MDL_002846147]; Ex. 589 [MSFT_AI_MDL_002881406]; Ex. 555 [MSFT_AI_MDL_002758540]; Ex. 590 [MSFT_AI_MDL_002758528]; Ex. 591 [MSFT_AI_MDL_002758535]; Ex. 592 [MSFT_AI_MDL_002758533]; *see also* Malackowski	662. **Disputed – Immaterial.** Microsoft disputes that it generated approximately [redacted] in revenue relating to the models and products at issue. This calculation is disputed to the extent it includes non-US revenue and revenue related to models and products not at issue in this case. Microsoft also disputes any implication that each and all of these revenue streams are reasonably related to the alleged infringement and reserves the right to challenge any damages calculation put forth by Plaintiffs on those grounds should this case proceed to trial. *On Davis*, 246 F.3d at 160 ("the term "gross revenue" under

NYT-DNP Opening Report at 150 (Figure 37, summarizing evidence) and Appendix D, Ex. 5.0.	the statute means gross revenue reasonably related to the infringement, not unrelated revenues"); News Ex. 127 (Rao Rebuttal Rep.) ¶¶ 9 (at pp. 13-14), 117-157. Microsoft also disputes Paragraph 662 as immaterial because these facts do not affect the outcome of the suit under the governing law. Paragraph 662 appears to relate entirely to Plaintiffs' damages claims. OpenAI defers to Microsoft's response.
663. From January 2023 to December 2024, Microsoft received revenue of approximately [redacted] in connection with its OpenAI revenue share for OpenAI API revenues and ChatGPT revenues. Ex. 538 [MSFT_AI_MDL_002758534] at Tab 1 ("RevShare to MSFT"); *see also* Malackowski NYT-DNP Opening Rep. at 146 and Appendix D, Exhibit 5.1.1.	663. **Disputed in Part – Immaterial.** Microsoft does not dispute that Microsoft received [redacted] in worldwide revenues from OpenAI related to OpenAI's API and ChatGPT revenue from January 2023 through December 2024. But Microsoft disputes that all of this revenue relates to the models and products at issue. This calculation is disputed to the extent it includes non-US revenue and revenue related to models and products not at issue in this case. Microsoft also disputes any implication that each and all of these revenue streams are reasonably related to the alleged infringement and reserves the right to challenge any damages calculation put forth by Plaintiffs on those grounds should this case proceed to trial. *On Davis*, 246 F.3d at 160 ("the term "gross revenue" under the statute means gross revenue reasonably related to the infringement, not unrelated revenues"); News Ex. 127 (Rao Rebuttal Rep.) ¶¶117 (at p.80), 118-143, 148-152, Tab 5c. Microsoft also disputes Paragraph 663 as immaterial because these facts do not affect the outcome of the suit

	under the governing law. Paragraph 663 appears to relate entirely to Plaintiffs' damages claims. OpenAI defers to Microsoft's response.
664. For the period of 2023 through 2024 Microsoft received revenue of [redacted] for OpenAI's use of the Microsoft Azure computing infrastructure to perform OpenAI operations—i.e., inferencing. Ex. 587 [MSFT_AI_MDL_002881404]; Ex 593 [MSFT_AI_MDL_000000972] at Slide 5; Ex. 601 [Griffith Depo. Tr.] at 53:11–54:14.; *see also* Malackowski NYT-DNP Opening Report at p. 146 and Appendix D, Exhibit 5.1.1.	664. **Disputed – Not Supported by Admissible Evidence; Immaterial.** Microsoft disputes that it received [redacted] in worldwide revenues from OpenAI related to inferencing from for the period 2023 through 2024. News Ex. 587 states that this revenue [redacted] News Ex. 587 (MSFT_AI_MDL_002881404). Microsoft disputes that all of this revenue relates to the models and products at issue. This calculation is disputed to the extent it includes non-US revenue and revenue related to models and products not at issue in this case. Microsoft also disputes any implication that these revenue streams are reasonably related to the alleged infringement and reserves the right to challenge any damages calculation put forth by Plaintiffs on those grounds should this case proceed to trial. *On Davis*, 246 F.3d at 160 ("the term "gross revenue" under the statute means gross revenue reasonably related to the infringement, not unrelated revenues"); News Ex. 127 (Rao Rebuttal Rep.) ¶¶ 117 (at p. 79), 118-152, Tab 5b. Microsoft additionally disputes Paragraph 664 to the extent it relies on Ex. 593 because it is not supported by

	admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Specifically, the cited evidence is hearsay and lacks foundation. Microsoft's 30(b)(6) witness on financials testified that she did not know who put this presentation together and "ha[d]n't seen this document before[,]" and referred to the financial data produced in the case as the source for financial data on the revenue share. Ex. 661 (K Griffith 30(b)(6) Tr.) A. Blanco Opp. Decl. Ex. 209 (K. Griffith 30(b)(6) Tr.) 52:20-53:9, 56:22-57:3, 63:5-18, 78:4-82:14. Microsoft also disputes Paragraph 664 as immaterial because these facts do not affect the outcome of the suit under the governing law. Paragraph 664 appears to relate entirely to Plaintiffs' damages claims. OpenAI defers to Microsoft's response.
665. Under the parties' agreements, Microsoft offers access to and use of OpenAI's LLMs to third parties through Microsoft API. Ex. 602 [MSFT_AI_MDL_000000575]; *see also* Ex. 601 [Griffith Depo. Tr.] at 28:9–12.	665. **Undisputed.**
666. Between January 2023 and December 2024, Microsoft generated net revenues of approximately [redacted] in connection with offering OpenAI's LLMs to third parties. Ex. 538 [MSFT_AI_MDL_002758534] at Tab "RevShare to OpenAI"; *see also* Malackowski NYT-DNP Opening Report at p. 147 and Appendix D, Exhibit 5.2.	666. **Disputed in Part – Not Supported by Evidence; Immaterial.** Microsoft does not dispute that it generated approximately [redacted] in worldwide revenue in connection with offering OpenAI's LLMs to third parties between January 2023 and December 2024. Microsoft disputes that all of this revenue relates to the models and products at issue. This calculation is disputed to the extent it includes non-US revenue and revenue

	related to models and products not at issue in this case. Microsoft also disputes any implication that these revenue streams are reasonably related to the alleged infringement and reserves the right to challenge any damages calculation put forth by Plaintiffs on those grounds should this case proceed to trial. *On Davis*, 246 F.3d at 160 (2d Cir. 2001) ("the term "gross revenue" under the statute means gross revenue reasonably related to the infringement, not unrelated revenues"); News Ex. 127 (Rao Rebuttal Rep.) ¶¶ 117 (at p. 75), 118-149, tabs 5a, 6a. Microsoft also disputes Paragraph 666 to the extent it relies on Ex. 538 because the cited evidence is not correct. Ex. 538 reflects revenue related to the revenue share agreement between Microsoft and OpenAI, not revenue related to Microsoft's offering OpenAI's LLMs to third parties. Microsoft also disputes Paragraph 666 as immaterial because these facts do not affect the outcome of the suit under the governing law. Paragraph 666 appears to relate entirely to Plaintiffs' damages claims. OpenAI defers to Microsoft's response.
667. Microsoft products that use OpenAI's models at issue include M365 Commercial, Security Copilot, Github Copilot, HCAm DAX Copilot, Copilot Pro, Copilot Freemium, M365 Copilot, Copilot for Service, Copilot for Sales, and Copilot Studio. Malackowski NYT-DNP Opening Report, pp. 148-149; Ex. 514 [Tinter 30(b)(6) Dep. Tr.] at	667. **Disputed in Part – Immaterial; Not Supported by Cited Evidence.** Microsoft does not dispute that the Microsoft products listed used an OpenAI model at issue in this case at some point during the period July 2022 through December 2024. But Microsoft disputes any implication that means all revenue for these products

52:20–53:6; Ex. 588 [MSFT_AI_MDL_002846147]; Ex. 603 [Griffith 30(b)(6) Depo. Tr.] at 105:3–106:18, 111:13–24, 173:10–23; Ex. 594 ["Microsoft 365," Microsoft, https://www.microsoft.com/en-us/licensing/product-licensing/microsoft-365]; Ex. 555 [MSFT_AI_MDL_002758540]; Ex. 604 [MSFT_AI_MDL_002758532]; Ex. 589 [MSFT_AI_MDL_002881406].	is reasonably related to the alleged infringement and reserves the right to challenge any damages calculation put forth by Plaintiffs on those grounds should this case proceed to trial. *On Davis*, 246 F.3d at 160 (2d Cir. 2001) ("the term "gross revenue" under the statute means gross revenue reasonably related to the infringement, not unrelated revenues"); News Ex. 127 (Rao Rebuttal Rep.) at ¶¶ 9 (at pp. 13-14), 117-157. Microsoft also disputes the remaining facts in paragraph 667 to the extent it makes representations about products after December 2024, because the cited evidence does not support any such representations. Additionally, Ex. 594 is a screenshot from a Microsoft website taken on September 3, 2026. Thus, the description of Microsoft products does not reflect the product at issue in this case during the relevant time period for which Microsoft produced financial data, July 2022 through December 2024. Additionally, Microsoft disputes the facts in Paragraph 667 as immaterial because these facts do not affect the outcome of the suit under the governing law. Paragraph 667 appears to relate entirely to Plaintiffs' damages claims. OpenAI defers to Microsoft's response.
668. According to invoices issued between April 2024 and December 2024, Microsoft generated revenues of approximately [redacted] in connection with selling its	668. **Undisputed.**

Bing Search API to OpenAI, allowing "OpenAI's model responses to be grounded in information from up-to-date Bing search results." Ex. 592 [MSFT_AI_MDL_002758533]; Malackowski NYT-DNP Opening Report p. 149, Appendix D, Exhibit 5.5.	

C. OpenAI and Microsoft's Relationship and Collaboration

1. Collaboration in Model Training

i) Microsoft's Provision of the Supercomputing Infrastructure that OpenAI Used to Train the At-Issue LLMs

669. OpenAI decided to partner with Microsoft because OpenAI needed computing support from Microsoft to train and develop its LLMs. Ex. 533 [Nadella Depo. Tr.] at 245:13-22 (early discussions in 2018 between OpenAI and Microsoft about a potential partnership were "driven by [OpenAI's] need for compute investments"; OpenAI needed to "either raise the money or get a partner to put up the money for a commercial arrangement."); Ex. 605 [Hoeschele 30(b)(6) Depo. Tr.] at 16:13-17 (OpenAI employee agreeing that "the reason that OpenAI entered its agreement with Microsoft for GPUs was because OpenAI needed GPUs to build, train, and operate its large language models").	669. **Disputed in Part - Not Supported by Admissible Evidence; Not Supported by Cited Evidence; Immaterial.** Defendants respond as follows: Undisputed that Hoeschele's testimony states that OpenAI needed GPUs to build, train, and operate its large language models and entered into an agreement with Microsoft. But, disputed to the extent News Plaintiffs assert the testimony of Mr. Nadella is admissible to establish that OpenAI "decided to partner with Microsoft because" it specifically needed computing support from Microsoft. The disputed proposition is not supported by admissible evidence, as required by Local Rule 56.1, and therefore is not properly before the Court. Mr. Nadella's testimony regarding OpenAI's intentions is hearsay. *See* A. Blanco Opp. Decl. Ex. 245 (S. Nadella 30(b)(1) Tr.) 245:13–22. Further, the disputed proposition is not supported by the cited evidence. Specifically, Hoeschele's testimony

	establishes that OpenAI needed GPUs and entered into an agreement with Microsoft for GPUs, but does not establish that OpenAI specifically needed Microsoft—as opposed to computing resources generally—or that OpenAI decided to partner with Microsoft solely for GPU support. *See* News Ex. 605 (P. Hoeschele 30(b)(6) Tr.) 16:13–17. Defendants also respond that OpenAI's need for computing support and that it chose Microsoft to supply that support is immaterial because these facts do not affect any claim or defense at issue in the current motions.
670. OpenAI could not have trained and developed the at-issue LLMs but for Microsoft's provision of computing resources to OpenAI. *See* Ex. 607 [Altman 30(b)(1) Depo. Tr.] at 198:3-9 ("Q. Why did you want to work with Microsoft. A. We weren't capable of building systems at this scale. And there were only a handful of companies in the world -- maybe Microsoft, Google, and Amazon at the time -- that could, and Microsoft was very value aligned with us in the way that the other companies were more skeptical of."); Ex. 595 [OPCO_MDL_004805725] at slide 4 (internal OpenAI presentation stating that "Compute is an area where OpenAI needs Microsoft's help"); Ex. 609 [Gaffney 30(b)(6) Depo. Tr.] at 202:21-203:12 (OpenAI employee confirming that, as of 2025, OpenAI had used "more than a gigawatt of compute in order to train, develop, research and serve its models to consumers," an essential component of its	670. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Defendants do not dispute that OpenAI needed computing support to train and develop its LLMs. But Defendants respond that the facts set forth in Paragraph 670 are disputed because they are not supported by the cited evidence; Microsoft was not the only entity capable of providing the needed support. Hurst Decl. Ex. 3 (S. Altman 30(b)(1) Tr.) 198:3-9 (listing Amazon and Google as potential partners). In particular, the cited evidence does not establish that OpenAI "could not have trained and developed" the at-issue LLMs without Microsoft's provision of computing resources. Altman testified that OpenAI itself was not capable of building systems at the required scale and identified multiple companies—including Microsoft, Google, and Amazon—that could provide such capabilities; his testimony does not establish that Microsoft specifically was necessary to developing the at-issue models. *See* Hurst Decl. Ex. 3 (S.

business because it "need[s] compute in order to train and serve our customers).	Altman 30(b)(1) Tr.) 198:3–9. Ex. 595 likewise states that compute was an area where OpenAI "needs Microsoft's help," but does not establish that Microsoft was the only possible source of the necessary computing resources. *See* Ex. 595 at OPCO_MDL_004805725, slide 4. And Gaffney's testimony establishes that compute is necessary to train and serve OpenAI's models, not that Microsoft's provision of compute was necessary or a but-for cause of OpenAI's ability to train and develop the at-issue LLMs. *See* Ex. A. Blanco Opp. Decl. Ex. 201 (Gaffney 30(b)(6) Tr.) 202:21–203:12. Defendants also respond that OpenAI's need for computing support and that it chose Microsoft to supply that support is immaterial because these facts do not affect the outcome of the suit under the governing law.
671. Microsoft knew that OpenAI was going to use Microsoft's supercomputer system to train large language models, and the system was designed specifically for OpenAI to train its LLMs. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 42:11-24, 89:13-24 (discussing the supercomputer Microsoft developed and admitting that "it was designed specifically to train OAI's AI models").	671. **Disputed in Part – Immaterial.** Microsoft does not dispute that it knew OpenAI was going to use Microsoft's supercomputer system to train large language models. But Microsoft disputes the characterization in Paragraph 671 that the systems were "designed specifically to train OAI's AI models" because it is not supported by the cited evidence. Mr. Melton testified that the design of the supercomputers Microsoft built for OpenAI "is the same design I use for every customer. It's just a bigger version of it," and that "[i]t wasn't a specific design for OpenAI's solution. . ." A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 87:3-15; *see also id.* 34:5-14. Melton agreed that the special design aspects were limited to meeting "the

	GPU count and speed requirements of OpenAI." *Id.* 36:8-15. Melton confirmed Microsoft did not know how OpenAI was going to train GPT models and "didn't know what [OpenAI was] going to use it for" or "how much was going to be used for what requirements." *Id.* 61:15-62:3. Microsoft further responds that Paragraph 671 is immaterial. Microsoft's alleged "knowledge" is not an element of any claim or defense at issue in News Plaintiffs' summary judgment motion. OpenAI defers to Microsoft's response.
672. Microsoft in May 2019 began work on the first phase of providing compute for training OpenAI's LLMs, including by collaborating with OpenAI on the physical and performance requirements for supercomputing systems that would be used to train OpenAI's LLMs. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 22:12-23:16.	672. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft does not dispute that in May 2019 it began work on the first phase of the supercomputer it was to provide to OpenAI. But Microsoft disputes the characterization in Paragraph 672 that the systems were "designed specifically to train OAI's AI models" because it is not supported by the cited evidence. Mr. Melton testified that the design of the supercomputers Microsoft built for OpenAI "is the same design I use for every customer. It's just a bigger version of it," and that "[i]t wasn't a specific design for OpenAI's solution. . ." A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 87:3-15; *see also id.* 34:5-14. Melton agreed that the special design aspects were limited to meeting "the GPU count and speed requirements of OpenAI." *Id.* 36:8-15. Melton confirmed Microsoft did not know how OpenAI was going to train GPT models and "didn't know what

	[OpenAI was] going to use it for" or "how much was going to be used for what requirements." *Id.* 61:15-62:3. Defendants further respond that the fact stated in Paragraph 672 is immaterial because this fact does not affect the outcome of the suit under the governing law. Whether Microsoft provided compute to OpenAI, some of which would be used to train OpenAI's LLMs, or collaborated regarding system requirements is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment. OpenAI defers to Microsoft's response.
673. These requirements for this supercomputing system were memorialized in a July 2019 agreement between Microsoft and OpenAI called the Joint Development and Collaboration Agreement. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 43:14-21; Ex. 596 [MSFT_AI_MDL_000792102] at -110.	673. **Undisputed – Immaterial.** Defendants respond that the fact stated in Paragraph 673 is immaterial because this fact does not affect the outcome of the suit under the governing law. The memorialization of requirements for the supercomputing system is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
674. Microsoft delivered the promised supercomputing system to OpenAI in May 2020, and OpenAI then began using the system. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 26:3-14, 28:7-12, 45:11-17.	674. **Undisputed - Immaterial.** Defendants respond that the fact stated in Paragraph 674 is immaterial because this fact does not affect the outcome of the suit under the governing law. The delivery and usage of the supercomputing system is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.

675. Microsoft's work on this supercomputing system was critical to OpenAI. Ex. 597 [OPCO_MDL_001937223] at -226, -227 (Sam Altman writing in September 2019 that "Our relationship with Microsoft is so important to our mission success," including because Microsoft has "the capability of delivering us, for the next 5 years at least, the most powerful supercomputers in the world").	675. **Disputed in Part - Not Supported by Cited Evidence; Immaterial.** Defendants respond as follows: Undisputed that News Ex. 597 contains the quoted statement. *See* News Ex. 597 at OPCO_MDL_001937226–27. However, disputed that the cited evidence establishes that Microsoft's "work on this supercomputing system" was "critical to OpenAI." The disputed proposition is not supported by the cited evidence. Specifically, News Ex. 597 does not state that Microsoft's work on any particular supercomputing system was "critical" to OpenAI. Defendants further respond that the fact stated in Paragraph 675 is immaterial because this fact does not affect the outcome of the suit under the governing law. Whether Microsoft's "work on this supercomputing system" was "critical to OpenAI" is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
676. Microsoft in late 2019 and early 2020 began working on the second phase of providing computing resources to OpenAI, a phase code-named Telemachus. Ex. 611[Melton 30(b)(6) Depo. Tr.] at 28:13-17, 31:8-10.	676. **Undisputed - Immaterial.** Defendants respond that the fact stated in Paragraph 676 is immaterial because this fact does not affect the outcome of the suit under the governing law. Microsoft's provision of computing resources to OpenAI is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
677. Once again Microsoft collaborated with OpenAI to tailor its support to OpenAI's precise needs, including OpenAI's request for a much larger GPU count to train even "larger" models. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at	677. **Disputed in Part - Not Supported by Cited Evidence; Not Supported by Admissible Evidence; Immaterial.** Defendants respond as follows: Undisputed that News Ex. 611 contains the quoted

28:18-29:25 ("OpenAI gave us new requirements and we met those requirements….").	testimony. However, disputed to the extent Plaintiffs characterize Mr. Melton's testimony as establishing OpenAI's "precise needs." The disputed proposition is not supported by admissible evidence, as required by Local Rule 56.1, and therefore is not properly before the Court. Specifically, Melton testified as Microsoft's corporate representative; his testimony about OpenAI's internal assessment of its own "precise need[s]" is inadmissible hearsay and lacks foundation and personal knowledge. Further, the disputed proposition is not supported by the cited evidence. Mr. Melton testified that the design of the supercomputers Microsoft built for OpenAI "is the same design I use for every customer. It's just a bigger version of it," and that "[i]t wasn't a specific design for OpenAI's solution. . . .". A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 87:3-15. Melton agreed that the special design aspects were limited to meeting "the GPU count and speed requirements of OpenAI." *Id.* 36:8-15. Defendants further respond that the fact stated in Paragraph 677 is immaterial because this fact does not affect the outcome of the suit under the governing law. That Microsoft provided storage space, some of which OpenAI used for training, is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
678. OpenAI's requirements for this phase of the project (Telemachus) are set forth in the March 5, 2021	678. **Undisputed - Immaterial.** Defendants respond that the fact stated in Paragraph 678 is immaterial

Amended and Restated Joint Development and Collaboration Agreement between Microsoft and OpenAI. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 48:14-49:5; Ex. 599 [MSFT_AI_MDL_000004071] at -084, -085, -101.	because this fact does not affect the outcome of the suit under the governing law. The requirements for Telemachus are not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
679. Microsoft delivered the Telemachus supercomputing resources to OpenAI in phases, delivering the first 18,000 GPUs around July 2, 2021. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 31:11-24, 53:25-54:18; Ex. 600 [MSFT_AI_MDL_000037384] at -384 ([redacted]	679. **Undisputed - Immaterial.** Defendants respond that the fact stated in Paragraph 679 is immaterial because this fact does not affect the outcome of the suit under the governing law. The delivery and size of Telemachus is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
680. Microsoft, in late 2024, completed the third phase of its project, codenamed Englewood, to provide supercomputing resources to OpenAI and released the results of that work to OpenAI. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 34:12-21.	680. **Undisputed - Immaterial.** Defendants respond that the fact stated in Paragraph 680 is immaterial because this fact does not affect the outcome of the suit under the governing law. The provision and release of supercomputing resources to OpenAI is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
681. Englewood was [redacted] than Telemachus in terms of GPU count. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 84:15-18.	681. **Undisputed - Immaterial.** Defendants respond that the fact stated in Paragraph 681 is immaterial because this fact does not affect the outcome of the suit under the governing law. The size of Englewood is not

	relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
682. GPT-3 was trained using Phase 1 (Owl) of Microsoft's Azure supercomputing system, and GPT-3.5 and GPT-3.5 Turbo were trained using Phase 2 (Telemachus). Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 40:5-15; 84:1-13.	682. **Disputed in Part - Not Supported by Cited Evidence; Not Supported by Admissible Evidence; Immaterial.** Defendants respond as follows: Disputed to the extent Plaintiffs rely on Melton's testimony to establish how OpenAI trained GPT-3, GPT-3.5, and GPT-3.5 Turbo. The disputed proposition is not supported by admissible evidence, as required by Local Rule 56.1, and therefore is not properly before the Court. Specifically, Melton testified as Microsoft's, not OpenAI's, Rule 30(b)(6) representative, and he is not authorized to testify on OpenAI's behalf concerning those activities; his testimony therefore lacks foundation as to how OpenAI trained those models. *See* A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 40:5–15, 84:1–13. Further, the disputed proposition is not supported by the cited evidence. Defendants also respond that the fact stated in Paragraph 682 is immaterial because this fact does not affect the outcome of the suit under the governing law. The fact that OpenAI used Microsoft's Azure supercomputing system is irrelevant to any asserted claim or defense at issue in News Plaintiffs' summary judgment motion.
683. All of OpenAI's GPT models from GPT-3 onward were trained using Microsoft's Azure supercomputing infrastructure. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 41:4-11.	683. **Disputed in Part - Not Supported by Cited Evidence; Not Supported by Admissible Evidence; Immaterial.** Defendants respond as follows: Disputed to the extent Plaintiffs rely on Melton's testimony

	to establish how OpenAI trained GPT models from GPT-3 onward. The disputed proposition is not supported by admissible evidence, as required by Local Rule 56.1, and therefore is not properly before the Court. Specifically, Melton testified as Microsoft's, not OpenAI's, Rule 30(b)(6) representative, and he is not authorized to testify on OpenAI's behalf concerning those activities; his testimony therefore lacks foundation as to how OpenAI trained those models. *See* A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 40:5–15, 84:1–13. Further, the disputed proposition is not supported by the cited evidence. Defendants also respond that the fact stated in Paragraph 683 is immaterial because this fact does not affect the outcome of the suit under the governing law. The fact that OpenAI used Azure infrastructure is irrelevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
684. Microsoft also provided OpenAI with multiple petabytes of Blob Storage—physical spinning medial-like discs storing big blocks of data—which OpenAI used to store training data. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 57:7-62:25.	684. **Disputed in Part – Not Supported by Cited Evidence; Not Supported by Admissible Evidence; Immaterial.** Defendants do not dispute that Microsoft provided Blob Storage to OpenAI. But Microsoft disputes the characterization in Paragraph 684 that Blob Storage was provided "for training data" because it is not supported by the cited evidence. Mr. Melton testified that Blob Storage "is not a AI thing. It's a generic storage solution" that "is used for many things including checkpoints, data, OS." A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 57:7-20, 61:15–62:3.

	Further disputed by OpenAI to the extent Plaintiffs rely on Melton's testimony to establish how OpenAI stored training data. The disputed proposition is not supported by admissible evidence, as required by Local Rule 56.1, and therefore is not properly before the Court. Specifically, Melton testified as Microsoft's, not OpenAI's, Rule 30(b)(6) representative, and he is not authorized to testify on OpenAI's behalf concerning those activities; his testimony therefore lacks foundation as to how OpenAI stored training data. *See* A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 57:7–62:25. Defendants further respond that the fact stated in Paragraph 684 is immaterial because this fact does not affect the outcome of the suit under the governing law. That Microsoft provided storage space, some of which OpenAI used for training, is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
685. That storage was critical to training the at-issue LLMs. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 57:7-62:25; Ex. 606 [MSFT_AI_MDL_000208547] at -548 [redacted]").	685. **Disputed in Part – Not Supported by Cited Evidence; Not Supported by Admissible Evidence; Immaterial.** Defendants do not dispute that OpenAI used Azure infrastructure to train and deploy GPT-3 and later models. Defendants do not dispute that storage space is needed to store any form of data, including training data, and that Microsoft provided OpenAI with storage space (called Blob) to utilize as needed. But Plaintiffs' cited evidence does not establish that the identified storage was "critical" to training the at-issue LLMs. Pursuant to Fed. R.

	Evid. 106, Melton testified that Blob storage "is critical for any training," and "it's not specific to OpenAI, that statement." *See* A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 61:4-63:7. Defendants further respond that the fact stated in Paragraph 685 is immaterial because this fact does not affect the outcome of the suit under the governing law. Whether Microsoft-provided storage space was "critical" to OpenAI's training is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment. OpenAI further disputes the facts set forth in Paragraph 685 to the extent Plaintiffs attribute these statements to OpenAI. Mr. Melton testified as Microsoft's Rule 30(b)(6) representative, not OpenAI's, and his designation by Microsoft does not authorize him to testify on OpenAI's behalf. *See* Fed. R. Civ. P. 30(b)(6). Likewise, the statement in News Ex. 606 was made by a Microsoft employee, not OpenAI, and does not constitute a representation by OpenAI concerning its training activities.
686. Microsoft also provided Azure supercomputing systems and resources for OpenAI to perform OpenAI operations, i.e. inferencing (e.g., operating the models to generate outputs), and make their models available to users, in the project code-named Project Mumford. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 85:14-87:2.	686. **Undisputed – Immaterial.** Defendants do not dispute that Microsoft provided Azure supercomputing systems and resources for OpenAI to perform OpenAI operations, i.e. inferencing (e.g., operating the models to generate outputs), and make their models available to users. However, any implication that Microsoft's provision of infrastructure "for OpenAI to

	perform operations" is disputed to the extent it implies a but-for causal relationship—i.e., that OpenAI could not have performed operations such as inferencing or made its models available to users without Microsoft's specific systems. Such an implication is not supported by the cited evidence. Defendants further respond that the fact stated in Paragraph 686 is immaterial because this fact does not affect the outcome of the suit under the governing law. That Microsoft "provided Azure supercomputing systems and resources for OpenAI to perform OpenAI operations…and make their models available to users" is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
687. This means that when OpenAI makes ChatGPT available to users, OpenAI is using Microsoft's Azure systems to do that. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 85:14-87:2.	687. **Undisputed – Immaterial.** Defendants do not dispute that OpenAI is using Microsoft's Azure systems when it makes ChatGPT available to users. Defendants further respond that the fact stated in Paragraph 687 is immaterial because this fact does not affect the outcome of the suit under the governing law. That OpenAI "is using Microsoft's Azure systems" when it "makes ChatGPT available to users" is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
688. Under its agreements with OpenAI, Microsoft has maintained the sole discretion, ability, and right—with or without cause—to revoke OpenAI's access to Azure altogether, i.e. to revoke its quota of resources and prevent	688. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Defendants respond as follows: Undisputed that Mr. Melton testified that Microsoft has a policy "for any customer" under which it

OpenAI from having access to the GPUs OpenAI was using to train and power its LLMs. Ex. 611 [Melton 30(b)(6) Depo. Tr.] at 92:10-98:6; Ex. 608 [MSFT_AI_MDL_000169501] at -502 (Statement of Work Agreement titled "Supercomputer Tech Collateral Exchange with OpenAI", executed June 30, 2020, which provides: "Microsoft reserves the right to suspend or revoke access to the Supercomputing Tech, at any time with or without cause.").	may revoke quota, or that the cited Statement of Work provides that "Microsoft reserves the right to suspend or revoke access to the Supercomputing Tech, at any time with or without cause." *See* News Ex. 608 at MSFT_AI_MDL_000169502. However, disputed that this right extended to Azure, as Plaintiffs assert. The cited contract language refers only to "Supercomputing Tech," and the assertion in Paragraph 688 expands the scope of the contractual provision beyond its terms. Further, the characterization of this right as one of "sole discretion" is disputed because that term does not appear in the cited contract language. The right to "suspend or revoke" access may also be subject to other contractual provisions—including notice requirements, implied covenants of good faith and fair dealing, and other terms of the parties' agreements—that Paragraph 688 omits. Further, Mr. Melton testified that he had not seen the cited Statement of Work and that he "was speculating and giving you generic company policy." A. Blanco Opp. Decl. Ex. 238 (L. Melton 30(b)(6) Tr.) 94:24-95:9. Defendants further respond that the fact stated in Paragraph 688 is immaterial because this fact does not affect the outcome of the suit under the governing law. Whether Microsoft maintained any such "discretion, availability, and right" is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.

ii) Microsoft's Assistance in Securing Additional Training Data for OpenAI

689. Microsoft uses a bot ("Bingbot") to crawl the internet and index and collect content from webpages, including News Plaintiffs' webpages, and Microsoft then stores that content in the "Bing Index." Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 21:17-22:6; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 63:12-64:2.	689. **Disputed in Part – Not Supported by Cited Evidence.** While Microsoft does not dispute that it may crawl and index content from News Plaintiffs' webpages to the extent permitted by News Plaintiffs' robots.txt and/or meta tag directives, Microsoft responds that the facts set forth in Paragraph 689 are disputed in part because the cited portions of News Exhibits 613 and 615 do not state that Bingbot crawls and collects content from "News Plaintiffs' webpages." *See* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) 21:17-22:6; A. Blanco Opp. Decl. Ex. 204 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) 63:12-64:2. OpenAI defers to Microsoft's response.
690. The Bing Index "supports the Bing search engine by crawling, storing, and organizing the internet." Ex. 610 [Sept. 25, 2025 Discovery Hearing Tr.] at 69:20-21.	690. **Undisputed - Not Supported by Admissible Evidence.** Microsoft objects and disputes the purported "fact" set forth in Paragraph 690 on the grounds that Mr. Briant is not a witness in this proceeding. *Flores v. Bondi*, 152 F.4th 73, 81 (2d Cir. 2025) (unsworn attorney statements are not evidence); *Sterling Select II Advisory, LLC v. Argus Info. And Advisory Servs., Inc.*, 2026 WL 734047 at *9 (S.D.N.Y. Mar. 16, 2026) (same); N.Y. Rule Prof. Conduct 3.7(a) (prohibition on lawyer as advocate). Arguments at a Discovery Hearing are not judicial admissions, nor are judicial admissions evidence. *In re Motors Liquidation Co.*, 957 F.3d 357, 360 (2d Cir. 2020); *Lovati v. Petroleos de Venezuala, S.A.*, 2025 WL 3677013 at *2 (S.D.N.Y. Dec. 18, 2025). Plaintiffs are not

	entitled to force Mr. Briant to be a witness by offering his prior arguments as evidence. *See Scottsdale Ins. Co. v. McGrath,* 2024 WQL 4512210 at *5-6 (S.D.N.Y. 2024) (excluding prior in-court arguments of counsel pursuant to Rule 403). Without waiving any of the foregoing, Microsoft does not dispute the accuracy of the statements quoted in Paragraph 690, and further responds that there is also ample evidence in the record that similarly describes how the Bing Index functions. *See, e.g.*, Gaggara Decl. ¶¶ 7-29. OpenAI defers to Microsoft's response.
691. According to Microsoft's counsel, the Bing Index "churns" through at least "billions of URLs a day." Ex. 610 [Sept. 25, 2025 Discovery Hearing Tr.] at 70:7-9.	691. **Undisputed – Not Supported by Admissible Evidence.** For the reasons identified in response to Paragraph 690, the Court should disregard the statement in Paragraph 691 in its entirety. Without waiving any of the foregoing, Microsoft does not dispute the accuracy of the statements quoted in Paragraph 690, and further responds that there is also ample evidence in the record that similarly describes how the Bing Index functions. *See, e.g.*, Gaggara Decl. ¶¶ 7-29. OpenAI defers to Microsoft's response.
692. [redacted]. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 224:13-228:8; Ex. 617[Canel	692. **Disputed in Part – Not Supported by Cited Evidence.** Microsoft does not dispute that [redacted]

30(b)(6) Depo. Tr.] at 36:12-24, 41:15-43:18; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 68:25-69:9.	[redacted] URLs. *See* ECF 1705 (P. Gaggara Decl.) ¶ 17. But Microsoft disputes the claims in Paragraph 692 that the Bing Index contains all content that may (or may not) exist for every URL indexed. If Bingbot discovers a URL that is selected for indexing, the contents of that webpage (if any) may be indexed only according to the permissions or restrictions communicated through robots.txt directives and/or meta tags. ECF 1705 (P. Gaggara Decl.) ¶¶ 10-18; ECF 1711 (F. Canel Decl.) ¶¶ 8-12; A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr.) 150:6-153:12, 157:16-24. Sometimes, the Bing Index will store a URL but no content associated with that webpage when consistent with the robots.txt and/or meta tag directives applied to that webpage. ECF 1705 (Gaggara Decl.) ¶¶ 20-23; A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr.) 68:20-69:5. OpenAI defers to Microsoft's response.
693. On November 3, 2019, Microsoft and OpenAI executed a "Statement of Work # 1 – Bing Index Data," through which Microsoft agreed to share a copy of the Bing Index with OpenAI. Ex. 612 [MSFT_AI_MDL_000003514] [redacted]	693. **Disputed in Part.** Defendants do not dispute that Microsoft and OpenAI entered into the Statement of Work #1-Bing Index Data on November 6, 2019. But Defendants dispute the statement in Paragraph 693 that "Microsoft agreed to share a copy of the Bing Index" as a mischaracterization of the terms of the agreement. Pursuant to that agreement, Microsoft [redacted] News Ex. 612 at MSFT_AI_MDL_000003514 (emphasis added). Additionally, the parties to the agreement agreed Microsoft

	would make such data available to OpenAI "solely for use by OpenAI and LP for 'computational analysis' (defined [redacted]). " *See* Ex. 612; *see also* Hurst Decl. Ex. 38 (P. Gaggara Tr. (Aug. 25, 2025) (30(b)(6)), at 93:5-95:2 (explaining that the term "computational analysis" means [redacted]).
694. Microsoft provided OpenAI with the full offline version of the Bing Index, and OpenAI in exchange paid Microsoft [redacted]. Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 145:25-146:7; Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 97:12-98:24; Ex. 610 [Sept 25, 2025 Hearing Tr.] at 76:17-77:3.	694. **Disputed in Part – Incomplete.** Defendants do not dispute that Microsoft provided portions of the Bing Index to OpenAI. But Defendants dispute the assertion in Paragraph 694 that Microsoft provided OpenAI "the full offline version of the Bing Index." The offline version of the Bing Index that Microsoft provided to OpenAI excluded webpages applying the NOARCHIVE and/or NOCACHE meta tag directives, and Microsoft removed content subject to DMCA takedown protocols from the Bing Index data it shared with OpenAI. A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan 27, 2026)) 230:22-231:8 ("We removed anything which was 'noarchive/nocache' from them. We removed anything which was DMCA removed from them."); *see also id.* 237:4-13, 232:8-233:23; A. Blanco Opp. Decl. Ex. 251 (M. Parakhin 30(b)(1) Tr.) 91:5-94:21, 101:19-102:16. Ms. Gaggara explained: "So given the state that was being experimented for training, there is no point in using 'noarchive/nocache' data because it's the same value, and

	it was not very useful. It was never going to be a part of the corpus even in future is how we interpreted the data sharing. And, hence, it was removed." Gaggara Jan. 27, 2026 30(b)(6) Tr.) at 233:7-23. Defendants also dispute any implication in Paragraph 694 that OpenAI's payments were "in exchange" for the Bing Index or that Microsoft profited from the arrangement, as it is not supported by the cited evidence. The payments Microsoft received from OpenAI in connection with Microsoft's provision of Bing Index data were reimbursement for costs incurred through maintaining the compute and infrastructure. Pursuant to Fed. R. Evid. 106, Ms. Gaggara's full testimony is: "Yeah. This was for the service -- the crawler service that we were running. That is what they paid for, because we were hosting the -- their service, and there was a compute cost. There was a temporary storage cost associated with it." A. Blanco Opp. Decl. Ex. 204 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) 146:3-7; *see also* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. Jan. 27, 2026) at 92:23-93:4 ("[OpenAI] does not pay. They were covering for the expenses for the hardware and any developmental costs involved."); *id.* 183:2-24 ("They paid for the costs that we incurred for the crawling service. . . . There was no revenue recognized."). OpenAI further objects on foundation grounds to Plaintiffs' reliance on Microsoft witnesses' testimony

	concerning OpenAI's conduct absent a sufficient showing of personal knowledge.
695. Microsoft's provision of the Bing Index was codenamed "Project Taxi." Ex. 414 [Sohl 30(b)(6) Dec. 9, 2025 Depo. Tr.] at 344:19-24(describing "Project Taxi" as the "internal OpenAI code name for a collaboration with Microsoft to evaluate the contents of web-crawled data used for Bing for the purposes of training an[] LLM model"); Ex. 521 [Yuan Depo. Tr.] at 77:18-21 (admitting that "Project Taxi" referred to the "project of evaluating the data from Bing crawl and evaluating the possibility of using it to train our large language model"); Ex. 522 [OPCO_MDL_004405641] at -641.	695. **Undisputed.**
696. "Project Taxi" was "considered highly confidential" within OpenAI. Ex. 523 [Passos Depo. Tr.] at 203:2-204:9.	696. **Disputed in Part – Immaterial**. OpenAI responds as follows: Undisputed that "Project Taxi" was considered confidential. Disputed that whether an internal project was considered highly confidential is relevant to the claims or defenses at issue, and OpenAI objects under Fed. R. Evid. 401 and 403. Further disputed as immaterial for the same reason and because this fact does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
697. After receiving payment for the Bing Index, Microsoft began providing OpenAI with a "few billion document[s]" taken from the Bing Index data in late 2019. Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 36:12-24, 41:15-43:18;	697. **Disputed in Part – Not Supported by Cited Evidence.** Defendants do not dispute that News Ex. 617 contains the quoted language. But Defendants dispute the statement that Microsoft began transferring the sample

Ex. 612 [MSFT_AI_MDL_000003514] at -514 (executed "Statement of Work" agreement for "Bing Index Data", dated November 3, 2019).	set of data "[a]fter receiving payment" from OpenAI because this is not supported by the cited evidence and contradicted by other record evidence. OpenAI did not make any payments, whether for cost recovery or otherwise, for the Bing Index data sample Microsoft shared in late 2019. *See* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) 96:20-24 ("Q Okay. And had Microsoft billed OpenAI or received any payment from OpenAI associated with the earlier data drops from the Bing index that had occurred prior to January 2023?" "A Not that I'm aware of.").
698. Following negotiations that included OpenAI's Chief Technology Officer, Mira Murati, and Microsoft, the parties in October 2021 agreed that OpenAI would [redacted] Ex. 614 [OPCO_MDL_001833041] at -041.	698. **Disputed – Not Supported by Admissible Evidence.** Defendants dispute the facts set forth in Paragraph 698 as they are not supported by admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Specifically, the cited evidence is hearsay and lacks foundation as to Microsoft; it is a statement by an OpenAI employee as to actions being taken by OpenAI. Further disputed to the extent Plaintiffs suggest that all training datasets were represented in the referenced portion of the GPT-3 dataset. *See* News Ex. 662 at OPCO_MDL_001653588 (enumerating the specific subset of datasets OpenAI shared in November 2021). Additionally, Defendants further state that Microsoft made the Bing Index data available to OpenAI "solely for use by OpenAI and LP for 'computational analysis' (defined as [redacted]

	[redacted] *See* News Ex. 612 at MSFT_AI_MDL_000003514; *see also* Ex. 204 (P. Gaggara 30(b)(6) Tr. (Aug. 25, 2025)) 93:5-95:2 (explaining that the term "computational analysis" means [redacted]).
699. In October 2021, OpenAI's Chief Research Officer Bob McGrew described the deal as "a trade": "the idea here [was] that we're giving data to them and they are giving data to us." Ex. 616 [OPCO_MDL_001936752] at -752.	699. **Disputed in Part – Incomplete; Not Supported by Cited Evidence.** Defendants respond as follows: Undisputed that News Ex. 616 reflects that OpenAI was "[redacted]." However, disputed that McGrew affirmatively "described the deal as 'a trade.'" Rather, McGrew asked whether the contemplated data exchange would result in fewer restrictions "because it's a trade." News Ex. 616 at OPCO_MDL_001936752. Plaintiffs' selective quotation converts McGrew's question into an affirmative characterization of the parties' arrangement. Plaintiffs' characterization is therefore not supported by the cited evidence. Defendants further dispute any implication that Microsoft transferred the entire Bing Index to OpenAI. Microsoft's corporate designee testified that the Bing Index data that Microsoft provided to OpenAI excluded webpages applying the NOARCHIVE and/or NOCACHE meta tag directives, and Microsoft removed any web content subject to DMCA takedown protocols. A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan 27, 2026)) 230:22-231:8

	("We removed anything which was 'noarchive/nocache' from them. We removed anything which was DMCA removed from them.").
700. Microsoft in late 2021 began providing larger portions of the Bing Index to OpenAI on a recurring basis, and by mid-2022 Microsoft had provided the full Bing Index to OpenAI, including [redacted] documents from webpages. Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 69:1-9, 106:5-21, 113:14-114:9, 1191:4-120:20; Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 59:13-25.	700. **Disputed in Part – Incomplete.** Defendants do not dispute that Microsoft began sharing samples of Bing Index data in later 2021. But Defendants dispute that "by mid-2022 Microsoft had provided the full Bing Index to OpenAI." While News Ex. 613 at 59:13-25 reflects Ms. Gaggara's testimony that the offline version of the Bing Index was shared with OpenAI by mid-2022, pursuant to Fed. R. Evid. 106, Ms. Gaggara later corrected her testimony and clarified that Microsoft shared the offline Bing Index with OpenAI in early 2023. A. Blanco Opp. Decl. Ex. 202 (P.Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) 229:13-230:10; *see also* A. Blanco Opp. Decl. Ex. 204 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) 69:1-9 (confirming offline Bing Index was shared in 2023). Defendants further respond that the facts set forth in Paragraph 700 are disputed as to the claim that "Microsoft had provided the full Bing Index to OpenAI." The portion of Bing Index data that Microsoft provided to OpenAI excluded webpages using the NOARCHIVE and/or NOCACHE meta tag directives, and Microsoft removed content subject to DMCA takedown protocols from the Bing Index data it shared with OpenAI. A. Blanco Opp. Decl. Ex. 202 (P.Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) 230:22-231:1 ("We removed anything which was 'noarchive/nocache' from them. We removed anything which was DMCA

	removed from them."); *see also id.* at 237:4-13, 230:8-233:23; A. Blanco Opp. Decl. Ex. 251 (M. Parakhin 30(b)(1) Tr.) 91:5-94:21. Ms. Gaggara explained: "So given the state that was being experimented for training, there is no point in using 'noarchive/nocache' data because it's the same value, and it was not very useful. It was never going to be a part of the corpus even in future is how we interpreted the data sharing. And, hence, it was removed." A. Blanco Opp. Decl. Ex. 202 (P.Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) 233:7-23. Additionally, Microsoft further states that Microsoft made the Bing Index data available to OpenAI "solely for use by OpenAI and LP for 'computational analysis' (defined as [redacted])." *See* News Ex. 612; *see also* A. Blanco Opp. Decl. Ex. 204 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) 93:5-95:2 (explaining that the term "computational analysis" means [redacted]).
701. Because of "[redacted]. Ex. 618 [MSFT_AI_MDL_000255338] at -340.	701. **Disputed in Part – Incomplete.** Defendants do not dispute that Exhibit 618 contains the quoted language in two separate sentences, but Defendants dispute any implication in Paragraph 701 of a causal connection between Microsoft's provision of Bing Index data on the receipt of GPT-3 training data. Pursuant to Fed. R. Evid. 106, Karmel Allison testified that "[t]here's no implication of causality" between the Microsoft's provision

	of portions of Bing Index data and OpenAI's provision of portions of the GPT-3 training data to Microsoft. A. Blanco Opp. Decl. Ex. 162 (K. Allison 30(b)(1) Tr. (Nov. 14, 2025)) at 70:24-72:8. Further disputed to the extent Plaintiffs imply that OpenAI shared all of GPT-3 training data with Microsoft. OpenAI shared access to three GPT-3 training datasets [redacted] in November 2021. This partial access did not include all of the training data for GPT-3. Further, the disputed proposition is not supported by the cited evidence. Microsoft further disputes the facts set forth in Paragraph 701 to the extent News Plaintiffs attribute these statements to Microsoft. The statements in News Ex. 618 were made by an OpenAI employee, not Microsoft, and do not constitute a representation by Microsoft.
702. A former Microsoft employee, Karmel Allison, in September 2021 described the exchange of data between Microsoft and OpenAI as "horse-trading." Ex. 620 [MSFT_AI_MDL_000096524] at –529.	702. **Disputed in Part – Not Supported by Cited Evidence; Immaterial; Incomplete.** Defendants do not dispute that Ex. 620 contains the quoted statement. However, Defendants dispute that Allison "described the exchange of data between Microsoft and OpenAI as 'horse-trading.'" The disputed proposition is not supported by the cited evidence. Specifically, the cited statement refers to potential negotiations [redacted] not to the exchange of data between Microsoft and

	OpenAI generally. Allison was discussing [redacted] News Ex. 620 at –MSFT_AI_MDL_000096529. Additionally, pursuant to Fed. R. Evid. 106, Karmel Allison testified: "I feel like you're overusing that characterization [horse trading] now" and that "[t]here's no implication of causality" between the Bing Index work and OpenAI's provision of portions of the GPT-3 training data to Microsoft. A. Blanco Opp. Decl. Ex. 162 (K. Allison 30(b)(1) Tr. (Nov. 14, 2025)) at 70:24-72:8. Further disputed as immaterial because a former Microsoft employee's description of an exchange of data between Microsoft and OpenAI does not affect the outcome of the suit under the governing law. Specifically, how a former Microsoft employee "described" an event is not a fact about an event and is therefore immaterial to the issues before the Court.
703. The Bing Index that Microsoft provided to OpenAI included News Plaintiffs' content. Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 68:25-69:13; Ex. 235 [MSFT_AI_MDL_000229826] at -27; Ex. 513 [Ribas Tr.] 174:17-19; Ex. 654 [MSFT_AI_MDL_001798013] at -15 ([redacted]	703. **Disputed – Not Supported by Cited Evidence.** Defendants respond that the facts set forth in Paragraph 703 are disputed because they are not supported by the cited evidence. The evidence cited as News Ex. 615 does not state that the Bing Index samples provided to OpenAI included News Plaintiffs' content, and the evidence cited as News Exs. 235, 513, and 654 do not discuss the Bing Index data that Microsoft provided to OpenAI.

[redacted]	Microsoft further disputes that it provided News Plaintiffs' paywalled content in the Bing Index to OpenAI. The portion of Bing Index data that Microsoft provided to OpenAI excluded webpages using the NOARCHIVE and/or NOCACHE meta tag directives, which web publishers use to indicate that their content is paywalled. A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) 237:4-13, 230:8-233:23; A. Blanco Opp. Decl. Ex. 251 (M. Parakhin 30(b)(1) Tr.) 91:5-94:21; *see also* ECF 1711 (F. Canel Decl.) ¶¶ 11(c) ("NOARCHIVE and NOCACHE meta tags have since evolved into a way for web publishers to indicate that their content is subscription-based or paywalled."), 29 ("[T]he vast majority of web publishers already used NOARCHIVE and/or NOCACHE to indicate paywalled or subscription-based web content ...."). Ms. Gaggara explained: "So given the state that was being experimented for training, there is no point in using 'noarchive/nocache' data because it's the same value, and it was not very useful. It was never going to be a part of the corpus even in future is how we interpreted the data sharing. And, hence, it was removed." A. Blanco Opp. Decl. Ex. (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 233:1-23 (emphasis added). Additionally, Defendants further state that Microsoft made the Bing Index data available to OpenAI "solely for use by OpenAI and LP for 'computational analysis' (defined as [redacted] *See*

	News Ex. 612 at MSFT_AI_MDL_000003514; *see also* A. Blanco Opp. Decl. 204 (P. Gaggara 30(b)(6) Tr. (Aug. 25, 2025)) 93:4-95:2 (explaining that the term "computational analysis" means [redacted]).
704. Microsoft employee Karmel Allison [redacted] Ex. 618 [MSFT_AI_MDL_000255338] at -341.	704. **Disputed in Part – Incomplete; Immaterial.** Defendants do not dispute that the cited document contains the quoted text. But Defendants dispute Paragraph 704 because it does not reflect any discussion concerning data "beyond the GPT-3 training data." Further, Defendants dispute any implication in Paragraph 704 that Microsoft conditioned its provision of Bing Index data to OpenAI on the receipt of a subset of GPT-3 training data from OpenAI. Pursuant to Fed. R. Evid. 106, Karmel Allison testified that "[t]here's no implication of causality" between the Bing Index work and OpenAI's provision of portions of the GPT-3 training data to Microsoft. A. Blanco Opp. Decl. Ex. 162 (K. Allison 30(b)(1) Tr. (Nov. 14, 2025)) at 70:24-72:8; *see also id.* at 69:20-23 ("Note that we don't actually need to fully pause things."). Defendants further dispute Paragraph 704 to the extent that it implicates "additional data beyond the GPT-3 training data" because this fact does not affect the outcome of the suit under the governing law. Specifically, Paragraph 704 does not state what [redacted] encompasses or how it is within the scope of the issues in this action.

705. OpenAI employee Katie Mayer in November 2021 described the "Bing data" as being "Mikhail's to sell," referring to Mikhail Parakhin, Microsoft's head of consumer AI. Ex. 622 [OPCO_MDL_002072434] at -435.	705. **Disputed in Part.** OpenAI responds as follows: Undisputed that News Ex. 622 contains the quoted statement. However, disputed to the extent Plaintiffs take Mayer's statement out of context. Pursuant to Fed. R. Evid. 106, Mayer testified that [redacted] A. Blanco Opp. Decl. Ex. 237 (K. Mayer 30(b)(1) Tr.) at 160:25–161:4. Microsoft joins.
706. Microsoft knew that OpenAI was in the process of developing commercial large language models and that OpenAI would use the Bing Index to further that work. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 44:15-45-13; Ex. 624 [MSFT_AI_MDL_000118608] at -609 (Microsoft's Chief Technology Officer Kevin Scott writing in [redacted]"); Ex. 626 [MSFT_AI_MDL_000046570] at -571 (Mr. Scott writing to Microsoft CEO Satya Nadella in [redacted]); Ex. 628 [MSFT_AI_MDL_000467526] at –526 (2023 Microsoft PowerPoint presentation describing [redacted]).	706. **Disputed in Part.** Microsoft does not dispute that Microsoft was aware that OpenAI was in the process of developing commercial large language models and provided Bing Index data to OpenAI to further its large language model development. Microsoft disputes any implication that Microsoft provided the Bing Index data for "developing **commercial** language models" and that OpenAI used the Bing Index data for training commercially released LLMs. Microsoft did not provide, and OpenAI did not use, the Bing Index data for pre-training OpenAI's production models. *See* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 147:18-148:11 ("Initially when the data was -- it was experimental purposes. It was not being used in any way, which was becoming a product of its own or it was not taking shape in any form. It was analysis."); A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 44:11-45:13 ("One, the Bing index data was never meant to be training

	material. It was only meant for experimentation. The Bing crawler -- the Mango crawler, which is what is being referred to here as 'Bing crawler,' was for generating training data."); A. Blanco Opp. Decl. Ex. 277 (I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) at 345:5-20 (OpenAI did not use Bing Index data in its "larger training regime for production models"). Microsoft also states that the claimed "July 2021" date for the evidence cited as News Ex. 626 is incorrect because News Ex. 626 at –MSFT_AI_MDL_000046571 is an email dated in July 2019. OpenAI defers to Microsoft's response.
707. Microsoft did not seek or obtain permission from any copyright holders whose works are included within the Bing Index before sharing the Bing Index with OpenAI. Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 38:10-17.	707. **Disputed in Part.** Microsoft does not dispute that Mr. Canel testified as follows during his deposition: "Q. Did Microsoft ask permission from any copyright holders, including authors and news organizations, about whether Microsoft could copy their copyrighted works and share them with third parties?" "No." But Microsoft does dispute any implication that it knowingly and/or willfully shared "copyrighted content" contained in the Bing Index in a manner inconsistent with responsible search engine practices. ECF 1705 (P. Gaggara Decl.)¶¶ 20-23 ("Microsoft takes measures to avoid including unauthorized copyrighted content in the Bing Index"); *see also* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 52:15-53:15. Microsoft removed content subject to DMCA takedown protocols from the Bing Index data it

	shared with OpenAI. A. Blanco Opp. Decl. Ex. 251 (M. Parakhin 30(b)(1) Tr. (Nov. 18, 2025)) at 101:19-102:16. Microsoft also filtered web content applying the NOARCHIVE or NOCACHE meta tags, regardless of whether such content was copyrighted. A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 237:4-9 ("Microsoft followed all the directives that were available to webmasters to filter the index content before passing it on. If there was news content which was not marked as 'noarchive'/'nocache,' it would have been there. If it was marked, it wouldn't be there."); *see also id.* at 230:22-231:1; A. Blanco Opp. Decl. Ex. 251 (M. Parakhin 30(b)(1) Tr. (Nov. 18, 2025)) at 91:5-94:21. OpenAI defers to Microsoft's response.
708. OpenAI used the Bing Index data provided by Microsoft to develop, test, and evaluate its large language models as well as training data for its large language models. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 66:5-68:25, 158:14-159:7; Ex. 414 [Sohl Dec. 9, 2025 30(b)(6) Depo. Tr.] at 345:5-13.	708. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI used the provided Bing Index data "to train several internal-only research models for evaluation purposes. A. Blanco Opp. Decl. Ex. 277 (I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) at 345:5-13. However, disputed to the extent Plaintiffs assert that OpenAI used the Bing Index data as training data for production models. The disputed proposition is not supported by the cited evidence. Specifically, OpenAI's corporate designee testified that OpenAI did not include the Bing Index data in its "larger training regime for production models," A. Blanco Opp. Decl. Ex. 277 (I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) at 345:14-20, and the only models trained on the

	Bing Index were "models used in evaluating [the Bing Index's] efficacy." A. Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr. (Jan. 29, 2026)) at 68:10-25. Microsoft joins.
709. In addition, Microsoft and OpenAI both use the Bing Index for retrieval augmented generation. Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 153:13-155:21; Facts 974, 992.	709. **Disputed in Part – Not Supported by Admissible Evidence; Not Supported by Cited Evidence.** Defendants do not dispute that Microsoft and OpenAI have both used Bing search for retrieval augmented generation. But Defendants respond that Paragraph 709 is not supported by admissible evidence, as required by Local Rule 56.1, and therefore is not properly before the Court. Specifically, News Ex. 617 is speculation and/or lacks foundation. Mr. Canel testified that, "based on the public statements of OpenAI," it was his "understanding [was] that they [OpenAI], indeed, are leveraging Bing." A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026) at 155:3-21. Defendants further dispute the claim in Paragraph 709 that OpenAI "use[s] the Bing Index for retrieval augmented generation" because this is not supported by the cited evidence. None of the cited evidence states that OpenAI uses the "Bing Index" for web search retrieval capabilities in its generative AI products. *See* A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 155:3-21 ("leveraging Bing"); ECF 1985-1 (News SF 974) (defining the "Bing API"); ECF 1985-1 (News SF 992) (citing to News Ex. 813 at slide 4 as supporting evidence,

	which states that ChatGPT Browsing "uses the Bing API"); *see also* A. Blanco Opp. Decl. Ex. 161 (E. Abib 30(b)(6) Tr.) at 68:24-71:21 (explaining differences between Microsoft's use of the Bing search engine endpoint for grounding and OpenAI's use of the Bing API for grounding, including that [redacted]). Adam Perelman also testified that Bing search was the first source of data for RAG, but OpenAI developed its own index at some point. A. Blanco Opp. Decl. Ex. 254 (A. Perelman 30(b)(6) Tr.) at 131:17-132:1 ("Q. Okay. And then at some point, OpenAI started using its own index, which is referred to as the Maraschino index. correct? A. Yes.").
710. Microsoft employees believed Microsoft violated search industry norms by supplying content that publishers had allowed it to crawl for search indexing to OpenAI as training data in the form of the Bing Index. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 146: 1-3 ("I'm saying it's not rightful for us to use the data for a different use case than what it was acquired for."); *id.* 145:16-21 ("We did not think it was the right thing to do given we acquired Bing data for search purposes, and we felt it was rightful to give the option to the webmasters to use the identifiers that are use case that it is being used for.'); Ex. 630 [MSFT_AI_MDL_000462444] at -44 ([redacted]").	710. **Disputed – Incomplete; Not Supported by Cited Evidence.** Microsoft does not dispute that Ms. Gaggara made the statements quoted in News Ex. 613 at her deposition. But, pursuant to Fed. R. Evid. 106, Ms. Gaggara testified: "Initially when the [Bing Index] data was -- it was experimental purposes. It was not being used in any way, which was becoming a product of its own or it was not taking shape in any form. It was analysis. It was value assessment, whereas, finally, when something is being used to create a product, then they'll -- the webmasters who own the data should decide if that should become a part of a product. So the bar is very different between the two. One is experimentally when -- to decide if we should go to the webmasters and do something about it. Second is when it's getting used, then they have -- they should have a say." *See*

	A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026) at 147:18-148:11. Microsoft further disputes any implication that Microsoft provided the Bing Index data to OpenAI for training production models. Microsoft made the Bing Index data available to OpenAI "solely for use by OpenAI and LP for 'computational analysis' (defined as [redacted] News Ex. 612 at MSFT_AI_MDL_000003514; *see also* A. Blanco Opp. Decl. Ex. 204 (P. Gaggara 30(bo(6) Tr. (Aug. 20, 2025)) at 93:4-95:2 (explaining that the term "computational analysis" means [redacted]); Blythe Opp. Decl. Ex. 49_(I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) at 345:10–13 (testifying that OpenAI used the Bing Index data to train "internal-only research models for evaluation purposes"); *see also id.* at 445:17–18 ("Project Taxi was used for training, but not training a production model."). OpenAI defers to Microsoft's response.
711. When ChatGPT first began using retrieval augmented generation to respond to user queries, the first source that ChatGPT used for retrieval augmented generation was the Bing Index. Ex. 409 [Perelman 30(b)(6) Depo. Tr.] at 131:17-22.	711. **Disputed in Part.** Undisputed that ChatGPT used Bing web search results when it first began using retrieval augmented generation. But disputed as to the assertion in Paragraph 711 that ChatGPT used "the Bing Index" when it first began using retrieval augmented generation. While Mr. Perelman agreed with counsel's

	characterization that ChatGPT used "the Bing Index" for retrieval augmented generation, pursuant to Fed. R. Evid. 106, Mr. Perelman independently testified that ChatGPT used the Bing API: "The first version of browsing in ChatGPT would be one of browsing launched in -- or announced in March of 2023 used the Bing API. A. Blanco Opp. Decl. Ex. 254 (A. Perelman 30(b)(6) Tr. (Jan. 30, 2026) at 30:1-3; *see also* Hurst Decl. Ex. 1 (E. Abib 30(b)(6). Tr. (Nov. 18, 2025)) at 68:24-71:21 (explaining differences between Microsoft's use of the Bing search engine endpoint for grounding and OpenAI's use of the Bing API for grounding, including that [redacted]).
712. After providing the Bing Index to OpenAI, Microsoft in summer 2023 embarked on "Project Mango", an effort to help OpenAI acquire as many documents as possible to train LLMs. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 140:9-13 (Microsoft employee describing Project Mango as "an effort to help OpenAI bootstrap [the] ability to collect as many of the documents as possible for their training—training the large language models").	712. **Disputed in Part.** Defendants do not dispute that Microsoft provided crawling services to OpenAI ("Project Mango") to collect web content for OpenAI to use for training its LLMs under the terms of the parties' Web Crawling Services Agreement. News Ex. 634 at (OPCO_MDL_004771618). But Defendants dispute any implication in Paragraph 712 that the goal was to "acquire as many documents as possible" without limitation or regard to web content controls. Microsoft did not crawl for OpenAI webpages that used robots.txt directives indicating that the web publisher did not want their content crawled by GPTBot. A. Hurst Decl. Ex. 38 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) at 144:3-5; A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 96:21-97:10.

713. Under Project Mango, Microsoft crawled the Internet to "collect data for the purposes of OpenAI to train LLM models." Ex. 414 [Sohl Dec. 9, 2025 30(b)(6) Depo. Tr.] at 337:17-24.	713. **Undisputed.**
714. Under Project Mango OpenAI and Microsoft collaborated to scrape web content. Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 97:11-98:3; Ex. 613 [Gaggara Jan. 27, 2026 Depo. Tr.] at 140:9-13; Ex. 619 [Huynh Jan. 27, 2026 30(b)(6) Depo. Tr.] at 65:2-66:21; Ex. 632 [MSFT_AI_MDL_000458392] at -392 ("[redacted]").	714. **Disputed in Part – Not Supported by Cited Evidence.** Defendants do not dispute that OpenAI and Microsoft collaborated to perform a joint crawl of a subset of web domains under Project Mango under the terms of the parties' Web Crawling Services Agreement. News Ex. 634 (OPCO_MDL_004771618). But Microsoft disputes any implication from the phrase "scrape" that Microsoft retrieved and acquired on its own behalf the contents of the webpages the Mango crawler discovered. Under Project Mango, Microsoft provided crawling services by using its web crawler technology to discover URLs (i.e., addresses for specific webpages within a website) from a subset of domains, to the extent that website did not apply robots.txt directives disallowing the GPTBot user-agent from accessing that website's content. A. Hurst Decl.Ex. 38 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) at 144:3-5; A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 96:21-97:10. The web content crawled from accessible URLs (if any) was streamed directly to OpenAI's Azure storage container for OpenAI to process, which Microsoft could not access. A. Hurst Decl. Ex. 38 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) 145:4-9; A. Blanco Opp. Decl. Ex. 202 (P. Gaggara

	30(b)(6) Tr. (Jan. 27, 2026) 31:17-19, 144:3-5, 162:13-163:20.
715. Microsoft ran a service called the Mango crawler on behalf of OpenAI to crawl the internet using the Bing Index's URLs and provide the content to OpenAI to use for training. Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 143:17-144:15; Ex. 634 [OPCO_MDL_004771618] at -618.	715. **Disputed in Part.** Microsoft does not dispute that Microsoft provided crawling services to OpenAI ("Project Mango") to collect web content for OpenAI to use for training its LLMs under the terms of the parties' Web Crawling Services Agreement. News Ex. 634 (OPCO_MDL_004771618). But Microsoft disputes the claim that the Mango crawler would "crawl the internet using the Bing Index's URLs" to the extent it suggests that specific content was identified and selected to crawl from in connection with Microsoft's web crawling services. Microsoft provided crawling services by crawling from a subset of domains (i.e., the name of the website such as www.example.com) and, to the extent that website did not apply robots.txt directives disallowing the GPTBot user-agent from accessing that website's content, discover the URLs (i.e., addresses for specific webpages within a website) available within those domains. The domains from which the Mango crawler discovered content did not point to any specific content that may (or may not) have been contained in the webpages within those domains. Microsoft did not have visibility into the specific content or works OpenAI obtained from the domains from which Microsoft crawled content on OpenAI's behalf or from the domains from which OpenAI itself directly crawled content. *See* Ex. (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026))

	at 36:6-14 ("I don't have visibility into what they really crawled."). Microsoft further disputes that it crawled all "URLs" existing in the Bing Index without limitation. Microsoft (on behalf of OpenAI) and OpenAI itself each crawled web content from half of a subset of domains, not all domains that exist on the web. A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 34:16-38:13 ("Mango crawler was being run by Microsoft for OpenAI. … They would give a list, and that list of domains is what Microsoft crawled using that instance of crawler specific to OpenAI."); *see also* A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) 108:15 ("It's tiny versus the whole internet."). Additionally, Microsoft did not crawl webpages that used robots.txt directives indicating that the web Publisher did not GPTBot to crawl their content. A. Hurst Decl. Ex. 38 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) at 144:3-5; A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 96:21-97:10. Further, OpenAI was to use the data "[redacted]." Ex. 634 at OPCO_MDL_004771618. OpenAI joins.
716. The Microsoft Project Mango crawler was a "crawler crawling the web, per OpenAI request," "on behalf	716. **Undisputed.**

of OpenAI," and "externally disclosed as GPTBot." Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 97:11-98:3.	
717. The Mango crawler was created using the Bingbot code base. Ex. 619 [Huynh Jan. 27, 2026 30(b)(6) Depo. Tr.] at 65:2-66:21.	717. **Undisputed – Not Supported by Cited Evidence.** Microsoft responds that the facts set forth in Paragraph 717 are undisputed but that the Court should disregard this Paragraph because the cited deposition testimony does not relate to the Bingbot codebase used for the Mango crawler. OpenAI joins Microsoft's response.
718. Microsoft performed this work between August 2023 and July 2024. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 140:19-141:1; Ex. 636 [MSFT_AI_MDL_001051845] at -847 (August 1, 2023 internal Microsoft email stating "We have a go ahead" and "Let's get started").	718. **Undisputed.** OpenAI defers to Microsoft's response.
719. Project Mango's "crawling was performed on Microsoft servers and maintained by Microsoft employees." Ex. 414 [Sohl Dec. 9, 2025 30(b)(6) Depo. Tr.] at 338:8-13.	719. **Disputed in Part.** Microsoft does not dispute that News Ex. 414 contains the quoted language. But Microsoft disputes any implication that Microsoft had access to or supervisory authority over the content collected and in OpenAI's possession in connection with Project Mango. *See* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 36:6-14 ("I don't have visibility into what they really crawled."); *see also id.* at 143:23-145:11, 155:18-157:2. Microsoft further responds that the crawled content was streamed directly to OpenAI's Azure storage container, which Microsoft could not access.

	A. Hurst Decl. Ex. 38 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) at 145:4-9; A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 31:17-19. OpenAI defers to Microsoft's response.
720. OpenAI paid Microsoft [redacted] for Microsoft's role in Project Mango. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 197:13-19; Ex. 638 [MSFT_AI_MDL_000001011] at -014.	720. **Disputed in Part - Not Supported by Cited Evidence; Incomplete.** Defendants do not dispute that OpenAI reimbursed Microsoft for costs related to Project Mango. But Defendants dispute that the cited evidence states that OpenAI paid Microsoft [redacted] for Microsoft's role in Project Mango. The disputed proposition is not supported by the cited evidence. Specifically, Gaggara testified only that she "think[s] it was [redacted], if I remember right," expressly qualifying her recollection of the total amount paid. *See* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 197:13–19. News Ex. 638, reflects a payment schedule providing for [redacted] but does not establish that OpenAI ultimately paid Microsoft [redacted]. *See* News Ex. 638 at MSFT_AI_MDL_000001014. Microsoft responds that Paragraph 720 is disputed as to any implication that Microsoft profited from this agreement, as pursuant to Fed. R. Evid. 106, Ms. Gaggara testified: "[I]t is not revenue in layman terms because we are not making a profit here. It is pure cost." A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 198:17-199:2; *see also* A. Blanco Opp. Decl. Ex. 204 (P.

	Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) at 146:3-7 ("This was for the service -- the crawler service that we were running. That is what they paid for, because we were hosting the -- their service, and there was a compute cost. There was a temporary storage cost associated with it."); A. Blanco Opp. Decl. Ex. 277 (I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) at 343:24-344:1 ("[T]he agreement is more about the operational burden of running a web-crawling service rather than paying for content.").
721. Microsoft gave OpenAI a list of about [redacted] URLs or domains in the Bing Index to be crawled that would result in obtaining about [redacted] documents (matching the size of the offline Bing Index), and OpenAI identified domains on the list to be crawled by the Project Mango crawler. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 36:1-37:4, 188:22-190:22; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 145:14-19; Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 111:21-113:19; Ex. 640 [MSFT_AI_MDL_001052306] at -306.	721. **Disputed in Part – Incomplete.** Defendants do not dispute that Microsoft provided OpenAI a list of about [redacted] domains that could be crawled as part of Project Mango and OpenAI filtered out domains that it did not want to be crawled by the Mango crawler. But Defendants dispute any implication from the claim that Microsoft provided OpenAI a list of "URLs" that Microsoft provided OpenAI a list of specific web content to crawl. Microsoft provided OpenAI a list of domains (*i.e.*, the name of the website such as www.example.com), not URLs (*i.e.*, addresses for specific webpages within a website); the list of domains that Microsoft provided did not point to any specific works that may (or may not) have been contained in the webpages within those domains. *See* A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 112:17–24. Pursuant to Fed. R. Evid. 106, Mr. Canel testified, "You said that it's a long list of URL. This is not true, this is a list of domains." A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 114:23-24.

	Microsoft further responds that it did not have specific knowledge of content used by OpenAI, and regardless, Microsoft's alleged "knowledge" is not an element of any claim or defense at issue in News Plaintiffs' summary judgment motion. A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 36:6-14 ("I don't have visibility into what they really crawled."). Also pursuant to Fed. R. Evid. 106, News Ex. 640 states: [redacted]." News Ex. 640 at MSFT_AI_MDL_001052306.
722. OpenAI crawled the remaining domains (essentially half the internet) through a separate OpenAI crawler, which was developed with Microsoft's assistance. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 36:1-37:4, 188:24-190:22; 214:19-215:17; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 145:14-19; Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 111:21-113:19; Ex. 640 [MSFT_AI_MDL_001052306] at -306.	722. **Disputed in Part – Not Supported by Cited Evidence; Incomplete.** Defendants do not dispute that Microsoft and OpenAI each crawled about half of the subset of domains through separate crawlers. But Defendants dispute that OpenAI crawled "essentially half the internet" because this is not supported by the cited evidence. Microsoft (on behalf of OpenAI) and OpenAI itself each crawled web content from half of a subset of domains, not all domains that exist on the web. A. Hurst Decl. Ex. 38 (P. Gaggara 30(b)(6) Tr. (Aug. 25, 2025)) at 143:25-144:1 ("We would run a service on their behalf to crawl the URLs…."); A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 34:16-38:13; *see also* A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 108:13-15 ("It's tiny versus the whole internet.").

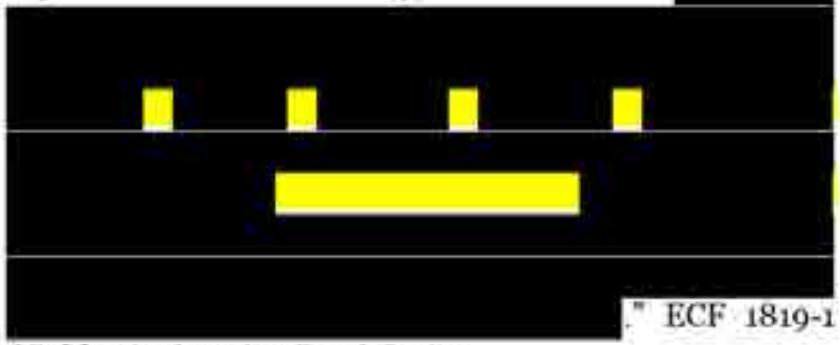
	Microsoft further disputes Paragraph 722 as follows: Microsoft disputes any implication that Microsoft developed OpenAI's independent crawler. Pursuant to Fed. R. Evid. 106, Ms. Gaggara testified: "Q And what assistance did Microsoft provide to OpenAI for operation of OpenAI crawler?" "A It was very self-sufficiently run by OpenAI team. Given they did not have prior knowledge of crawlers and what kind of issues they can run into, we would guide them in potential issues they may run into and what they should watch out for." A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 214:19-25.
723. Under "Project Mango," both Microsoft and OpenAI identified their crawlers as "GPTBot." Ex. 617 [Canel 30(b)(6) Deposition Tr.] at 97:11-98:15.	723. **Undisputed.**
724. [redacted] Ex. 632 [MSFT_AI_MDL_000458392] at -392_0003.	724. **Undisputed.**
725. [redacted]	725. **Undisputed.**

[redacted] Ex. 632 [MSFT_AI_MDL_000458392] at -392_0002.	
726. Microsoft intended to keep Project Mango "secret." Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 84:1-85:15; *see also* Ex. 642 [MSFT_AI_MDL_001053220] at –220 ("[redacted]").	726. **Disputed in Part – Incomplete.** Microsoft does not dispute that Ex. 624 contains the quoted language, but pursuant to Fed. R. Evid. 106, Microsoft states as follows: First, Microsoft was acting in its capacity as a technology service provider and did not disclose the activities of its customers consistent with typical business practices. A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 142:10-145:11 ("There was no attempt to keep it [Project Mango] secret or make it public. It was a technology we were offering OpenAI, and OpenAI was using it."). Second, to the extent websites had a complaint or questions regarding crawling activities, those questions would be properly directed toward OpenAI. A. Blanco Opp. Decl. Ex. 182 (Canel 30(b)(6) Tr.) at 85:25-86:19. OpenAI defers to Microsoft's response.
727. At least one Microsoft employee objected to Microsoft's role in Project Mango. Ex. 617 [Canel Depo. Tr.] at 12:16-13:6.	727. **Disputed in Part – Incomplete.** Microsoft responds as follows: Microsoft disputes any implication in Paragraph 727 that a Microsoft employee "objected to Microsoft's role in Project Mango" for any reason related to the issues in this case. Pursuant to Fed. R. Evid. 106, Mr. Canel testified that he "really would like to revisit the complexity of crawling" because "I'm very busy

	at Microsoft. This is an additional project for me. That's why every time there is a new project to make sure that we do appropriate prioritization of work." A. Blanco Opp. Decl. Ex. 182 (Canel 30(b)(6) Tr. (July 16, 2026)) at 77:3-25. OpenAI defers to Microsoft's response.
728. Microsoft's Project Mango crawler began operating in summer 2023 and ran for a year. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 140:9-141:11.	728. **Undisputed.**
729. This Project Mango data was delivered to OpenAI, and OpenAI used it to train LLMs, including specifically GPT-4.o, GPT-4.5, and GPT-5. Ex. 414 [Sohl Dec. 9, 2025 30(b)(6) Depo. Tr.] at 340:5-341:2; Ex. 621 [Ryder 30(b)(6) Tr.] at 162:9-22.	729. **Disputed in Part – Immaterial**. Defendants respond as follows: Undisputed that some data collected through Project Mango was delivered to OpenAI and used in training certain models. Disputed to the extent Plaintiffs' characterization suggests that all Project Mango data was used to train GPT-4o, GPT-4.5, and GPT-5. Sohl qualified his testimony, stating that "the data—or at least some part of the data collected under the GPTBot user agent—has been used in multiple models." A. Blanco Opp. Decl. Ex. 277 (I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) at 340:5–341:2. Ryder similarly testified that OpenAI combined GPTBot data and "used that data as the top of funnel for our web data in ongoing training runs," not that all data collected through Project Mango was incorporated into each identified model. A. Blanco Opp. Decl. Ex. 267 (N. Ryder 30(b)(6) Tr. (Jan. 29, 2026) at 162:9–22. Further, disputed to the extent training of GPT-4.5 and GPT-5 is immaterial because they are not at issue in this litigation.

730. The data Microsoft obtained from its Project Mango crawler was assembled into a dataset called [redacted] which contains 42,680 unique copies of The Times and Daily News Plaintiffs' articles. Goldstein Opening Rep. at ¶ 565.

730. **Disputed in Part – Incomplete; Expert Opinion Subject to Rule 702.** Defendants do not dispute that the [redacted] dataset was comprised of materials collected by both Microsoft on behalf of OpenAI and OpenAI itself in connection with Project Mango. But Defendants dispute any implication that web content from News Plaintiffs' domains was collected in connection with Project Mango. Pursuant to Fed. R. Evid. 106, Goldstein's full statement in his expert report is as follows: "Notably, the version of the [redacted]



." ECF 1819-1 (Goldstein Opening Rep.) ¶ 565.

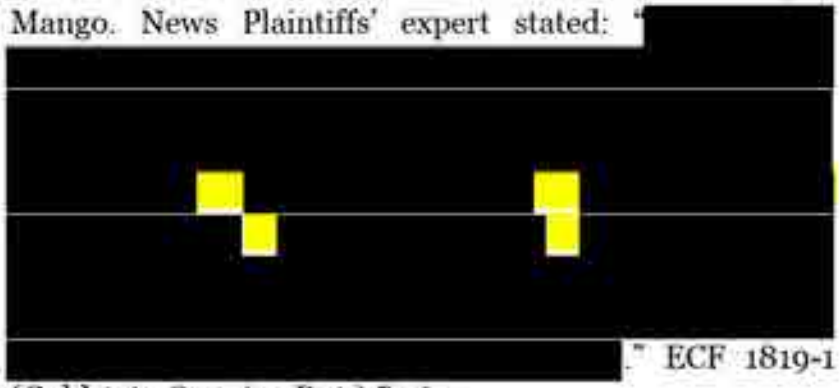
OpenAI further responds as follows: Disputed that the cited evidence establishes that the dataset contained "42,680 unique copies." That figure is the result of Goldstein's expert analysis and flawed methodology. OpenAI disputes Goldstein's flawed methodology for calculating "copies," as it lacks basic validation steps to ensure matching text includes protectable expression, calculate an error rate, or otherwise account for false positives, and has challenged it under

	Federal Rule of Evidence 702. The fact stated in Paragraph 730 is disputed because the evidence cited in support is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2).
731. The [redacted] contains 118,223 unique copies of Ziff Davis articles. [Delorey Opening Report] ¶¶ 217–219; *See* Delorey Supplemental Report ¶38; Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 74:3-12.	731. **Disputed - Expert Opinion Subject to Rule 702; Not Supported by Admissible Evidence.** OpenAI responds as follows: Disputed that the cited evidence establishes the existence of "118,223 unique copies." Delorey's analysis relied on a flawed Asserted Works corpus containing substantial third-party content—including press releases, advertising content, and content republished from other news outlets—and his unvalidated text-matching heuristics produced "confirmed matches" that plainly are not Ziff Davis copyrighted works, including a written statement authored by President Obama, an SEC press release, and sections of the United States Code. Pursuant to Fed. R. Evid. 106, *see* Defs.' Mem. of Law ISO Rule 702 Mot. at 23–24 (citing Malhotra Decl., Ex. 2 (Delorey Dep. 67:18–24, 106:13–14, 119:17–22, 268:9–11)). The fact stated in Paragraph 731 is disputed because the evidence cited in support is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Defendants have moved to exclude Delorey's "copies" opinion because his Asserted Works corpus contained substantial third-party content and he failed to

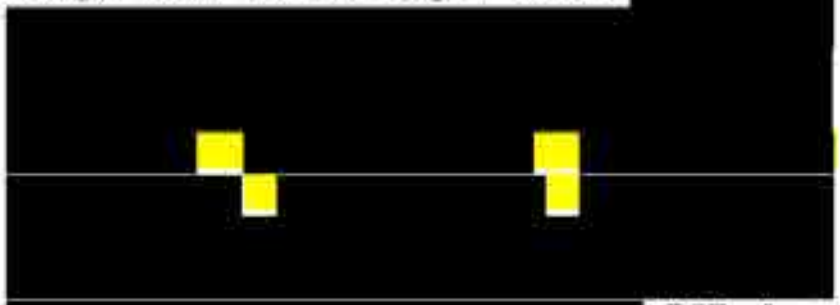
	validate his quality filters and text-matching. Pursuant to Fed. R. Evid. 106, *see* Defs.' Mem. of Law ISO Rule 702 Mot. at 23–24. *See* ECF 1774-1 (Delorey Opening Rep.) ¶¶ 217–219; ECF 1774-4 (Delorey Supplemental Rep.) ¶ 38. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). The facts set forth in Paragraph 731 are disputed because they are not supported by admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Microsoft responds that these facts do not pertain to claims asserted against Microsoft.
732. The dataset called [redacted] was train the GPT-4o model. Ex. 414 [Sohl Dec. 9, 2025 30(b)(6) Depo. Tr.] at 337:17-24, 341:6-342:11.	732. **Disputed - Not Supported by Cited Evidence.** Defendants respond as follows: Disputed as unsupported by the cited evidence. Specifically, the cited excerpts of deposition testimony do not state that the named dataset was used to train GPT-4o.
733. The Mango crawl data (i.e., [redacted]) includes News Plaintiffs' Asserted Works. *See infra* Section re: "Defendants' Infringement of News Plaintiffs' Asserted Works Ripe for Summary Judgment"	733. **Disputed - Not Supported by Cited Evidence.** Defendants respond as follows: Disputed as unsupported by any cited evidence. Disputed to the extent the statement suggests that the cited dataset contains all News Plaintiffs' Asserted Works at issue in this Motion. Further disputed to the extent the statements in the cross-referenced section are also disputed.
734. Before commencing Project Mango, Microsoft did not consult with the owners of the websites whose	734. **Disputed in Part.** Microsoft does not dispute the deposition testimony cited in Paragraph 734. But

content was crawled by the Mango crawler. Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 91:4-7.	Microsoft disputes that Microsoft did not communicate that it was crawling content on behalf of OpenAI under Project Mango in any form. Microsoft, as a party performing crawling services on behalf of OpenAI, used the GPTBot user-agent to identify the Mango crawler as OpenAI's crawler. *See* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Aug. 20, 2025)) 146:12-147:5; *see also* A. Blanco Opp. Decl. Ex. 182 (F. Canel Dep.30(b)(6) Tr.(Jan. 22, 2026)) at 78:5-9, 96:21-97:10. OpenAI publicly announced its GPTBot crawler, which News Plaintiffs were aware of and responded to. *See* News SUMF ¶ 484 ("In August 2023, OpenAI announced its bot called GPTBot which OpenAI and Microsoft have used to crawl the web for training data."), ¶ 488 ("The Times disallowed GPTBot in its robots.txt on August 17, 2023."), ¶ 495 ("Daily News Plaintiffs, except New York Daily News and Chicago Tribune, disallowed GPTBot in their robots.txt for their respective websites on August 21, 2023."), ¶ 496 ("New York Daily News disallowed GPTBot in its robots.txt for its website at least as of September 28, 2023."), ¶ 498 ("The Chicago Tribune disallowed GPTBot and ChatGPT-User in its robots.txt for its website at least as of April 30, 2024."). Additionally, Microsoft disputes the claim that it had begun crawling content before News Plaintiffs were made aware of OpenAI's GPTBot crawling activities. As stated above, OpenAI publicly announced its GPTBot crawler in August 2023. Microsoft began providing the content it crawled on behalf of OpenAI in September 2023. News Ex. 628 at –

	MSFT_AI_MDL_000467526_003 (Microsoft had begun providing crawled content to OpenAI in September 2023). OpenAI joins Microsoft's response.
735. Prior to Microsoft and OpenAI commencing Project Mango (in August 2023), The Times in April 2023 informed both Microsoft and OpenAI that they did not have The Times's permission to use The Times's content for training LLMs. Ex. 235 [MSFT_AI_MDL_000229826] at 830 (April 3, 2023 email from The Times to Microsoft); Ex. 139 [OPCO_MDL_004621506] at -07. (April 4, 2023 from The Times to OpenAI).	735. **Disputed in Part – Incomplete.** Defendants do not dispute that The Times communicated with Defendants in April 2023. Pursuant to Fed. R. Evid. 106, News Plaintiffs did not include Microsoft's response, which stated, in pertinent part, "Microsoft's primary interaction with New York Times ("NYT") content takes place as part of its operation of the Bing search engine, as with any other information published on the web. To provide relevant search results, Bing crawls online materials according to standard industry practice, including by honoring the robots.txt protocol. Bing does not crawl content from sites where a site owner has disallowed the Bing bot access. Webmasters can also control the content that can be accessed or cached by Bing via webmaster tools, more information on which can be found here. Microsoft also uses certain publicly available data ingested from the web for the purposes of training AI." News Ex. 235 at MSFT_AI_MDL_000229827. Defendants further state that web publishers communicate the permissions and restrictions on the use of their content through implementing robots.txt and meta tag directives. *See* ECF 1711 (F. Canel Decl. (Sept. 4, 2026)) ¶ 12 ("Because crawlers can only read the files, meta directives, and properties of a given webpage, rather than

	plain language, these longstanding content controls provide an easily implemented, universal way for web publishers to reliably communicate instructions to search engines."); *see also* ECF 1705 (P. Gaggara Decl. (Sept. 4, 2026)) ¶ 15. Defendants further dispute any implication in Paragraph 735 that Defendants collected content from News Plaintiffs' domains in connection with Project Mango. News Plaintiffs' expert stated: "." ECF 1819-1 (Goldstein Opening Rpt.) ¶ 565.
736. As Microsoft and OpenAI were scraping web content for Project Mango, The Times was engaged in discussions with each of Microsoft and OpenAI about deals to license Times content for training LLMs. Ex. 646 [NYT_00203287] at -87 (August 2023 email from The Times to OpenAI regarding "high level commercial terms"); Ex. 648 [MSFT_AI_MDL_002295146 at -49] (August 2023 email from The Times to Microsoft regarding "a high level overview of the partnership structure we have in mind for the use of NYT content").	736. **Disputed in Part – FRE 408; Not Supported by Cited Evidence.** Microsoft disputes the facts in Paragraph 736 because the cited evidence is not admissible because it concerns settlement negotiations subject to Fed. R. Evid. 408. Without waiving its FRE 408 objection, Microsoft joins in the response below. Defendants do not dispute that The Times and Defendants had discussions in the summer and fall of 2023. But Defendants dispute any implication in

Paragraph 736 that Defendants collected content from News Plaintiffs' domains in connection with Project Mango. News Plaintiffs' expert stated: [redacted]



[redacted] ECF 1819-1 (Goldstein Opening Rpt.) ¶ 565.

Defendants also dispute the claim that such communications occurred "[a]s Microsoft and OpenAI were scraping web content for Project Mango" because News Plaintiffs do not cite evidence to support this assertion. The communications cited as Exs. 646 and 648 occurred in August 2023, and, as stated in response to Paragraph 734, Microsoft began providing content it crawled on OpenAI's behalf under Project Mango beginning in September 2023. Defendants also dispute that "Microsoft ... scrap[ed] web content for Project Mango" because Microsoft performed crawling services, not "scraping" services.

Defendants further dispute Plaintiffs' characterization of the August 2023 communications as discussions about "deals to license Times content for training LLMs." Exhibit 646 reflects [redacted]

	[redacted] News Ex. 646 at NYT_00203287. Exhibit 648 reflects a [redacted] Ex. 648 at MSFT_AI_MDL_002295149. Neither exhibit reflects any substantive response, counterproposal, or agreement by Microsoft or OpenAI, and neither establishes that Microsoft or OpenAI agreed to "license Times content for training."
737. The Mango crawler did not respect publisher controls like the NOARCHIVE and NOCACHE meta tags; instead, Microsoft allowed OpenAI to decide what to respect and what not to respect. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 41:6-43:21, 162:13-164:9; Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 120:3-121:2; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 147:7-148:14; Ex. 638 [MSFT_AI_MDL_000001011] at -011 (October 16, 2023 agreement between OpenAI and Microsoft regarding "Web Crawling Assistance").	737. **Disputed in Part – Not Supported by Admissible Evidence.** Defendants respond that this fact is disputed because it is not supported by admissible evidence, as required by Local Rule 56.1, and therefore is not properly before the Court. Specifically, Ms. Gaggara, whose testimony is cited as Exhibits 613 and 617, testified that whether OpenAI committed to following meta tag directives would "be speculation on my behalf. I don't know what they did with it[.]" Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 41:6-43:13. Defendants further dispute Paragraph 737 as follows: Defendants dispute any implication that Microsoft could have, but chose not to, configure the Mango Crawler to recognize and filter NOARCHIVE/NOCACHE content.

	Web crawlers, like the Mango crawler, can only follow robots.txt directives; they do not have the technical ability to read meta tag directives applied to webpages. *See* Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 162:13-164:9 ("Mango crawler's responsibility was only to crawl content. And when you crawl content, all you have access to is robots or you have access to HTTP address. Unless you process the content, you cannot see what is 'noarchive' or 'nocache.'); *see also* A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 122:6-9. Defendants further dispute that the cited evidence establishes that Microsoft gave OpenAI discretion to disregard publisher controls. *See* News Ex. 638 at -012 (Agreement for "Web Crawling Assistance"); *see also* News Ex. 613 at 42:9-12. Further disputed to the extent Plaintiffs suggest that OpenAI chose to disregard such controls; the cited evidence in Paragraph 737 does not establish that OpenAI failed to honor applicable publisher controls during subsequent processing or use of the collected data.
738. Microsoft did not configure the Mango crawler to avoid crawling webpages with the NOARCHIVE meta tags, nor did it filter out those webpages before providing the crawled data to OpenAI. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 41:6-43:21, 162:13-164:9; Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 120:3-121:2; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 147:7-148:14; Ex. 638 [MSFT_AI_MDL_000001011] at -011 (October 16, 2023	738. **Disputed in Part.** Microsoft does not dispute that the Mango crawler did not avoid crawling webpages with NOARCHIVE meta tags from the data provided to OpenAI in connection with the Mango crawling services. But Microsoft disputes any implication that Microsoft could have, but chose not to, configure the Mango Crawler to recognize and filter NOARCHIVE/NOCACHE content. Web crawlers, like the

agreement between OpenAI and Microsoft regarding "Web Crawling Assistance").	Mango crawler, can only follow robots.txt directives; they do not have the technical ability to read meta tag directives applied to webpages. *See* A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026)) at 162:13-164:9 ("Mango crawler's responsibility was only to crawl content. And when you crawl content, all you have access to is robots or you have access to HTTP address. Unless you process the content, you cannot see what is 'noarchive' or 'nocache.'); *see also* A. Blanco Opp. Decl. Ex. 182 (F. Canel 30(b)(6) Tr. (Jan. 22, 2026)) at 122:6-9. Microsoft further incorporates its response to Paragraph 737 above. OpenAI defers to Microsoft's response.
739. OpenAI did not agree to respect the publishers' meta tags either. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 41:6-43:21, 162:13-164:9; Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 120:3-121:2; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 147:7-148:14; Ex. 638 [MSFT_AI_MDL_000001011] at -011 (October 16, 2023 agreement between OpenAI and Microsoft regarding "Web Crawling Assistance").	739. **Disputed in Part – Not Supported by Admissible Evidence; Not Supported by Cited Evidence.** Defendants respond that this fact is disputed because it is not supported by admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Specifically, Ms. Gaggara, whose testimony is cited as Exhibits 613 and 617, testified that whether OpenAI committed to following meta tag directives would "be speculation on my behalf. I don't know what they did with it[.]"A. Blanco Opp. Decl. Ex. 202 (P. Gaggara 30(b)(6) Tr. (Jan. 27, 2026) at 41:6-43:13. Defendants further incorporate their response to Paragraph 737 above.

	Defendants further dispute any implication in Paragraph 739 that meta tags are the only well-known mechanism by which website owners can express preferences regarding how their content may be accessed or used. *See* ECF 1711 (F. Canel Decl.) ¶¶ 10-12, 26-29; N Feamster Opp. Decl. ¶ 18.
740. The Mango crawler obtained content from paywalled news sites, which Microsoft transferred to OpenAI for use in training LLMs. Ex. 613 [Gaggara Jan. 27, 2026 30(b)(6) Depo. Tr.] at 41:6-43:21, 162:13-164; Ex. 617 [Canel 30(b)(6) Depo. Tr.] at 120:3-121:2; Ex. 615 [Gaggara Aug. 20, 2025 30(b)(6) Depo. Tr.] at 147:7-148:14; Ex. 638 [MSFT_AI_MDL_000001011] at -011 (October 16, 2023 agreement between OpenAI and Microsoft regarding "Web Crawling Assistance").	740. **Disputed – Not Supported by Cited Evidence.** Defendants respond that this fact is disputed because it is not supported by the cited evidence. News Ex. 638, the web crawling services agreement entered between Microsoft and OpenAI, does not state or suggest News Plaintiffs' content was crawled under Project Mango. News Ex. 638 at –012. Defendants further respond that Paragraph 740 is disputed because none of the cited evidence states that "[t]he Mango crawler obtained content from paywalled news sites." Instead, News Plaintiffs' expert stated, [redacted] Instead, the dataset includes copies of News Plaintiffs articles from numerous third-party sites." ECF 1819-1 (Goldstein Opening Rpt.) ¶ 565. Defendants further incorporate their response to Paragraph 737 above.

	Defendants further dispute any implication that crawlers are able to detect and respect client-side subscription enforcement controls which Plaintiffs characterize as "paywall[s]". Feamster Opp. Decl. ¶ ¶ 12-14.
741. Effective December 15, 2023, Microsoft and OpenAI also [redacted]. Ex. 650 [MSFT_AI_MDL_000003441] at -441.	741. **Disputed in Part – Not Supported by Cited Evidence.** Defendants respond as follows: Plaintiffs' cited evidence does not support the assertion. Specifically, Ex. 650 is a Statement of Work that addresses [redacted]
742. [redacted] Microsoft engineers worked on site at OpenAI to assist OpenAI in building and improving OpenAI's large language models. Ex. 650 [MSFT_AI_MDL_000003441] at -442.	742. **Disputed in Part - Not Supported by Cited Evidence.** Defendants respond as follows: Plaintiffs' cited evidence does not support the assertion. Specifically, Ex. 650 is a Statement of Work that [redacted] News Ex. 650 at – MSFT_AI_MDL_000003442. However, Ex. 650 does not [redacted] Defendants dispute any implication that Microsoft onsite engineers were involved in or had access to OpenAI's training data. Microsoft's corporate designee

	testified that [redacted] A. Blanco Opp. Decl. Ex. 214 (J. Huynh 30(b)(6) Tr. (Jan. 27, 2026)) at 282:19–23; 267:20–25. OpenAI's corporate designee likewise testified that [redacted] A. Blanco Opp. Decl. Ex. 267 (N. Ryder 30(b)(6) Tr. (Jan. 29, 2026)) at 168:1–11.

2. Financial Relationship Between Microsoft and OpenAI

743. [redacted] Ex. 652 [MSFT_AI_MDL_000010360] at -361.	743. **Undisputed – Immaterial.** Microsoft does not dispute that Kevin Scott made the statements in Ex. 652. But Microsoft responds that these statements are immaterial because they do not affect the outcome of the suit under the governing law. Paragraph 743 does not contain a fact of consequence to any claim or defense at issue in Plaintiffs' summary judgment motion. Microsoft further responds that the facts related to Google in Paragraph 743 are immaterial because Google's activities and capabilities are not relevant to the claims or defenses at issue. OpenAI defers to Microsoft's response.
744. [redacted]	744. **Undisputed.**

[redacted]." Ex. 652 [MSFT_ AI_MDL_000010360] at -361.	
745. [redacted]." Ex. 652 [MSFT_ AI_MDL_000010360] at -363.	745. **Undisputed.**
746. In July 2019, Microsoft and OpenAI entered into a partnership agreement, with Microsoft committing an initial $1 billion in capital to OpenAI. Ex. 574 [https://news.microsoft.com/source/2019/07/22/openai-forms-exclusive-computing-partnership-with-microsoft-to-build-new-azure-ai-supercomputing-technologies/]; Ex. 575 [MSFT_AI_MDL_000000345] (Joint Development and Collaboration Agreement); Ex. 576 [MSFT_AI_MDL_001550855] at -855.	746. **Undisputed.**
747. Under this 2019 agreement, Microsoft obtained the right to commercialize OpenAI's LLMs in Microsoft's products. Ex. 575 [MSFT_AI_MDL_000000345] (Joint Development and Collaboration Agreement); Ex. 656 [MSFT_AI_MDL_000040868] at -868; Ex. 533 [Nadella Depo. Tr.] 42:10-22.	747. **Disputed in Part.** Defendants respond as follows: Disputed to the extent Plaintiffs' assertion suggests that Microsoft obtained the right to commercialize all of OpenAI's LLMs in Microsoft products. Rather, the 2019 agreement granted Microsoft rights to commercialize certain OpenAI language-model technology in Microsoft's products, subject to the agreement's terms and limitations.

	News Ex. 575 at MSFT_AI_MDL_000000357–358, 361; News Ex. 656 at MSFT_AI_MDL_000040868.
748. Under a September 2020 amendment to the 2019 OpenAI-Microsoft agreement, OpenAI [redacted] Ex. 658 [MSFT_AI_MDL_000019962] at -967.	748. **Disputed in Part – Not Supported by Cited Evidence.** Defendants respond as follows: Undisputed that the September 2020 amendment states OpenAI [redacted] News Ex. 658 at MSFT_AI_MDL_000019967. However, disputed as to the remaining portion of Paragraph 748 because it is not supported by the record. Specifically, the cited evidence does not support the assertions that OpenAI [redacted] The cited evidence states that OpenAI [redacted] *See* News Ex. 658 at MSFT_AI_MDL_000019967. Indeed, Exhibit E expressly [redacted]. *Id.*

	Microsoft further disputes Paragraph 748 to the extent it implies that OpenAI actually provided Microsoft the [redacted]. The record shows that OpenAI provided Microsoft with only a subset of its GPT-3 training data in a tokenized format [redacted]. *See* A. Blanco Opp. Decl. Ex. 292 (B. Norick 30(b)(1) Tr.) at 47:3-11 (explaining that OpenAI transferred "[a] fraction of pretraining data"); A. Blanco Opp. Decl. Ex. 201 (B. Gaffney 30(b)(6) Tr.) at 148:2-6 (explaining that OpenAI granted Microsoft temporary access to a subset of GPT-3 training data); A. Blanco Opp. Decl. Ex. 278 (X. Song 30(b)(1) Tr. (Jan. 13, 2026)(b)(1)), at 82:23-83:13 ("One-third of the data was in the tokenized format, was transferred for the only purpose of evaluation, and [redacted]; A. Blanco Opp. Decl. Ex. 214 (J. Huynh 30(b)(6) Tr. (Jan. 27, 2026)(30b6) at 237:11-13 ("[T]he training datasets were tokenized, so not in natural language for a human to look at.").
749. OpenAI's corporate witness Benton Gaffney agreed that [redacted] Ex. 609 [Gaffney 30(b)(6) Depo. Tr.] at 145:7-146:4.	749. **Disputed in Part.** Defendants respond as follows: Undisputed that Gaffney answered "Yes" when asked whether the text of the 2020 amendment [redacted] OpenAI's counsel objected that the question was beyond the scope of Gaffney's Rule 30(b)(6) designation and called for a legal

	conclusion. A. Blanco Opp. Decl. Ex. 201 (B. Gaffney 30(b)(6) Tr.) at 145:7–146:4. Disputed to the extent Plaintiffs suggest that this testimony establishes [redacted], for the reasons set forth in Defendants' response to Fact No. 748. When subsequently asked whether the agreement [redacted] Gaffney did not affirm that interpretation, instead explaining that his understanding was that OpenAI had transferred only a subset of GPT-3 training data to Microsoft. *See id.* at 148:8–20; *see also* A. Blanco Opp. Decl. Ex. 281 (J. Tinter 30(b)(6) Tr.) at 237:17-239:10 ("[redacted]"). Defendants incorporate their response to Paragraph 748.
750. Microsoft accessed the GPT-3 training data in January 2021, as part of its "journey to commercialize OpenAI's state-of-the-art GPT-3 model in our 1P products and services." Ex. 660 [OPCO_MDL_004635421] at -421.	750. **Disputed in Part – Not Supported by Admissible Evidence.** Defendants respond as follows: Defendants do not dispute that in early 2021, a limited number of Microsoft engineers accessed a portion of OpenAI's GPT-3 training data. However, the remainder of Paragraph 750 is disputed. As Song testified, they did so [redacted]. Pursuant to Fed. R. Evid. 106, *see* A. Blanco Opp. Decl. Ex. 278 (X. Song 30(b)(1) Tr.) 82:25–83:13. And as Jamie Huynh, one of the purported authors listed on the draft document that is Ex. 660 testified, a small team

	at Microsoft received portions of the GPT-3 training data [redacted]. A. Blanco Opp. Decl. Ex. 214 (J. Huynh 30(b)(6) Tr.) at 189:11-17, 236:10-239:19. Further, Ex. 660 is a draft document and therefore cannot be substantiated as containing correct information and is thus not admissible evidence, as required by Local Rule 56.1. S.D.N.Y. Local Rule 56.1 (d). Indeed, as Mr. Gaffney testified, "I can't say, with 100 percent confidence, exactly what it's getting at." A. Blanco Opp. Decl. Ex. 201 (Gaffney 30(b)(6) Tr.) 150:12-25.
751. OpenAI also shared portions of the GPT-3 training data with Microsoft on at least one other occasion. Ex. 662 [OPCO_MDL_001653588] at -591.	751. **Undisputed.**
752. In March 2021, Microsoft and OpenAI entered into the Amended and Restated Joint Development and Collaboration Agreement. Ex. 599 [MSFT_AI_MDL_000004071] at -071.	752. **Undisputed.**
753. In connection with the March 2021 amended agreement, Microsoft committed to investing $2 billion in OpenAI, [redacted]. Ex. 599 [MSFT_AI_MDL_000004071] at -073; Ex. 623 [Wetter 30(b)(6) Depo. Tr.] at 29:5-30:24; Ex. 576 [MSFT_AI_MDL_001550855] at -855.	753. **Undisputed.** Microsoft does not dispute the facts stated in 753, but Microsoft responds that News Decl. Ex. 576 is a draft Microsoft memo that does not identify the author or recipient ("To:" and "From:" are blank), and therefore cannot be substantiated as containing correct information and is thus not admissible evidence, as required by Local Rule 56.1.

754. In January 2023, Microsoft and OpenAI entered into the Second Amended and Restated Joint Development and Collaboration Agreement. Ex. 602 [MSFT_AI_MDL_000000575] at -575.	754. **Undisputed.**
755. In connection with the January 2023 agreement, Microsoft committed to investing $10 billion in OpenAI, including $8 billion for supercomputing infrastructure and [redacted] Ex. 602 [MSFT_AI_MDL_000000575] at -592; Ex. 623 [Wetter 30(b)(6) Depo. Tr.] at 29:5-30:24; Ex. 576 [MSFT_AI_MDL_001550855] at -855.	755. **Undisputed.** Microsoft does not dispute the facts stated in 755, but Microsoft responds that News Decl. Ex. 576 is a draft Microsoft memo that does not identify the author or recipient ("To:" and "From:" are blank), and therefore cannot be substantiated as containing correct information and is thus not admissible evidence, as required by Local Rule 56.1.
756. As of 2023, Microsoft had committed to investing a total of $13 billion in OpenAI [redacted]." Ex. 576 [MSFT_AI_MDL_001550855] at -855.	756. **Undisputed.** Microsoft does not dispute the facts stated in 756, but Microsoft responds that News Decl. Ex. 576 is a draft Microsoft memo that does not identify the author or recipient ("To:" and "From:" are blank), and therefore cannot be substantiated as containing correct information and is thus not admissible evidence, as required by Local Rule 56.1.
757. By September 2025, $11.6 billion of Microsoft's $13 billion commitment to OpenAI had been funded. Ex. 664 [Microsoft Form 10-Q for the Quarterly Period Ended September 30, 2025] at 9.	757. **Undisputed.**

758. Under the January 2023 agreement, Microsoft is entitled to 20% of OpenAI's revenues. Ex. 602 [MSFT_AI_MDL_000000575] at -596; Ex. 533 [Nadella Depo. Tr.] at 131:21-132:11.	758. **Disputed in Part.** Defendants respond as follows: Defendants do not dispute that under the Second Amended JDCA, Microsoft was entitled to 20% of net revenue generated by OpenAI through commercial use of technology incorporating certain licensed intellectual property, subject to the agreement's terms. *See, e.g.*, News Ex. 602 at MSFT_AI_MDL_000000596. Defendants dispute that Microsoft receives 20 percent of OpenAI's total revenue, as the cited agreement provides that Microsoft would receive 20% of "Net Revenue generated by [OpenAI] during the Term through the Commercial Use of any Technology embodying any Licensed IP as permitted in [the] Agreement." News Ex. 602 at MSFT_AI_MDL_000000581 (defining "Net Revenue"), MSFT_AI_MDL_000000596. Microsoft further responds that revenue it receives from OpenAI is for only the OpenAI API and ChatGPT "[redacted] Additionally, [redacted]. A. Blanco Opp. Decl. Ex. 281 (J. Tinter 30(b)(6) Tr.) 210:23-212:12; 214:1-215:24; News Ex. 538 at MSFT_AI_MDL_002758534.
759. Microsoft's financial records indicate that, from January 2023 to December 2024, Microsoft received approximately [redacted] from OpenAI in connection with this revenue share. Ex. 538 [MSFT_AI_MDL_002758534].	759. **Undisputed – Immaterial.** This fact is repeated in Paragraph 663. Microsoft incorporates its response to Paragraph 663 above. OpenAI defers to Microsoft's response.

760. There is [redacted]." Ex. 514 [Tinter 30(b)(6) Depo. Tr.] at 313:13-22.	760. **Disputed.** Defendants respond as follows: The facts stated in Paragraph 760 are disputed. Microsoft's 30(b)(6) witness on financials testified that, while she "can't think of -- of a cost -- a cost of goods sold that would be specific to just receiving that rev share" that "there's other costs that Microsoft incurs for delivering Azure infrastructure to OpenAI so that OpenAI can generate their own revenue." A. Blanco Opp. Decl. Ex. 209 (K. Griffith 30(b)(6) Tr.) 55:24–56:10. Additionally, Microsoft witness Jon Tinter explained that [redacted] Plaintiffs' characterization does not take Microsoft's larger investment into account, but focuses on the Azure discount element only. A. Blanco Opp. Decl. Ex. 281 (J. Tinter 30(b)(6) Tr.) 313:13–314:6.
761. Microsoft receives additional revenue from OpenAI because OpenAI is using Microsoft's Azure computing infrastructure for inferencing and for delivery of services to OpenAI customers; this revenue stream is called Inferencing. Malackowski NYT-DNP Opening Rep. at 146; Ex. 593 [MSFT_AI_MDL_000000972] at slide 5.	761. **Disputed in Part – Not Supported by Cited Evidence; Not Supported by Admissible Evidence.** Microsoft does not dispute that it received revenue from OpenAI during the period July 2022 through December 2024 for "OpenAI's usage of Azure resources to support OpenAI's inferencing workloads" or that this revenue stream is called inferencing. *See* Ex. News 587 (MSFT_AI_MDL_002881404). Microsoft disputes the remaining facts in Paragraph 761 to the extent it purports to state facts about Microsoft revenue after December 2024 because the cited evidence does not support any such facts.

	Additionally, Microsoft disputes the facts in Paragraph 761 to the extent it relies on News Ex. 593 on the grounds that it is not supported by admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Specifically, the cited evidence is hearsay and lacks foundation. Microsoft's 30(b)(6) witness on financials testified that she did not know who put this presentation together and "ha[d]n't seen this document before[,]" and referred to the financial data produced in the case as the source for financial data on the revenue share. A. Blanco Opp. Decl. Ex. 209 (K. Griffith 30(b)(6) Tr.) 52:20-53:9, 56:22-57:3, 63:5-18. OpenAI defers to Microsoft's response.
762. Microsoft's financial records indicate that, from January 2023 to December 2024, Microsoft received approximately [redacted] from OpenAI in connection with Inferencing. Malackowski NYT-DNP Opening Rep. at 146, Appendix D, Exhibit 5.0; Ex. 593 [MSFT_AI_MDL_000000972] at slide 5.	762. **Disputed – Not Supported by Evidence; Not Supported by Admissible Evidence; Immaterial.** This fact is repeated in Paragraph 664. Microsoft incorporates its response to Paragraph 664 above. OpenAI defers to Microsoft's response.
763. Microsoft has generated additional revenue in connection with Azure OpenAI services, i.e., by offering OpenAI's LLMs to third parties through Azure. Malackowski NYT-DNP Opening Rep. at 147; Ex. 593 [MSFT_AI_MDL_000000972] at slide 6.	763. **Undisputed – Not Supported by Admissible Evidence.** Microsoft does not dispute that it has generated revenue during the period July 2022 through December 2024 in connection with Azure OpenAI services, which "enables customers to use OpenAI's models hosted in Azure." A. Blanco Opp. Decl. Ex. 209 (K. Griffith

	30(b)(6) Tr.) 28:9-12; News Ex. 590 (MSFT_AI_MDL_002758528) at Tab 1. Microsoft disputes the remaining facts in Paragraph 763 to the extent it purports to state facts about revenue after December 2024 because the cited evidence does not support any such facts. Microsoft further responds that, to the extent Paragraph 763 relies on News Ex. 593, it is not supported by admissible evidence, as required by Local Rule 56.1, and therefore are not properly before the Court. Specifically, the cited evidence is hearsay and lacks foundation. Microsoft's 30(b)(6) witness on financials testified that she did not know who put this presentation together and "ha[d]n't seen this document before[,]" and referred to the financial data produced in the case as the source for financial data on the revenue share. A. Blanco Opp. Decl. Ex. 209 (K. Griffith 30(b)(6) Tr.) 52:20-53:9, 56:22-57:3, 63:5-18. OpenAI defers to Microsoft's response.
764. Microsoft's financial records indicate that, from January 2023 to December 2024, Microsoft earned approximately [redacted] in connection with these services, 20% of which Microsoft must share with OpenAI under the parties' revenue sharing agreement. Malackowski NYT-DNP Opening Rep. at 147; Ex. 538 [MSFT_AI_MDL_002758534] at Tab "RevShare to OpenAI."	764. **Disputed in Part – Not Supported by Evidence; Immaterial.** This fact repeats the same facts set forth in Paragraph 666. Microsoft incorporates its response to Paragraph 666 above. Further disputed to the extent Plaintiffs suggest Microsoft shared 20% of all revenue in connection with Azure OpenAI services; the revenue share calculation accounts for certain discounts and other adjustments. Ex. 538

	[MSFT_AI_MDL_002758534] at Tab "RevShare to OpenAI." OpenAI defers to Microsoft's response.
765. As of September 2025, Microsoft held an approximately 27 percent equity stake in OpenAI's Public Benefit Corporation. Malackowski NYT-DNP Opening Rep. at 82; Ex. 664 [Microsoft Form 10-Q for the Quarterly Period Ended September 30, 2025] at 27.	765. **Disputed in Part – Not Supported by Cited Evidence.** Undisputed that as of October 2025, "Microsoft held approximately 27 percent in the PBC on an as-converted diluted basis." News Ex. 664 at 27. OpenAI further responds that this fact is disputed as to Plaintiffs' reference to September 2025. OpenAI's Public Benefit Corporation was not formally announced until October 28, 2025 and the referenced recapitalization agreement was executed on the same day. News Ex. 564; News Ex. 664 at 27.
766. Given OpenAI's $500 billion valuation as of October 2025, Microsoft's 27% stake was valued at approximately $135 billion. Ex. 664 [Microsoft Form 10-Q for the Quarterly Period Ended September 30, 2025] at 27.	766. **Disputed in Part – Not Supported by Cited Evidence.** Defendants do not dispute that as of October 2025, "Microsoft held approximately 27 percent in the PBC on an as-converted diluted basis." News Ex. 664 at 27. But Defendants respond that OpenAI's $500 Billion valuation is disputed because the cited evidence does not support the proposition.
767. Given OpenAI's $852 billion valuation as of April 2026, Microsoft's 27% stake was valued at approximately $230 billion. Ex. 584 [https://openai.com/index/accelerating-the-next-phase-	767. **Disputed in Part**. Defendants do not dispute OpenAI's $852 billion valuation as of April 2026, or Microsoft's 27% stake in the PBC on an "as-converted diluted basis." News Ex. 664 at 27. Disputed to the extent that the value of Microsoft's equity holdings in OpenAI is

ai/]; Ex. 664 [Microsoft Form 10-Q for the Quarterly Period Ended September 30, 2025] at 27.	an unrealized gain. Until Microsoft sells any of its equity, there is substantial uncertainty as to what returns may ultimately exist. Even if the current equity valuation were relevant to matters currently before the Court, it would be speculative to quantify what Microsoft's ultimate benefit may be. Thus, the disputed proposition is not supported by the cited evidence.

3. Microsoft's Ability to Stop OpenAI's Infringement

768. Under the Defendants' Joint Development and Collaboration Agreement ("JDCA"), Microsoft could require OpenAI to respond to "requests to provide additional detail or answer specific questions regarding Models or projects disclosed in the Initial Disclosure or any Subsequent Disclosure (e.g., specifically what data was used to train any given Model)." Ex. 602 [MSFT_AI_MDL_000000575] at –586; *see also* Ex. 533 [Nadella Depo. Tr.] at 59:15-60:11 (Microsoft's CEO acknowledging at deposition that Microsoft had "at least a monthly opportunity to ask OpenAI about the data it trained on" via a "governing board").	768. **Disputed in Part – Immaterial.** Defendants respond as follows: Undisputed that the Second Amended JDCA merely provides that "OpCo and OpenAI will respond to Microsoft's questions regarding Licensed IP (including requests to provide additional detail or to answer specific questions regarding Models or projects disclosed in the Initial Disclosure or any Subsequent Disclosure (e.g., specifically what data was used to train any given Model))." Disputed as to the scope of the parties' information-sharing agreement. The Second Amended JDCA itself, in the same paragraph plaintiffs cite, states that "[t]he Initial Disclosure and each Subsequent Disclosure are not intended to provide an exhaustive listing of all Licensed IP in or underlying any given Model or development project." News Ex. 602 at MSFT_AI_MDL_000000586. Rather, each disclosure "is intended to provide to Microsoft a high-level catalog of Models and development projects …, and a *brief description of each*, at the time of such disclosure." *Id.*

	(emphasis added). Nadella further testified that Microsoft did not have access to all the training data. A. Blanco Opp. Decl. Ex. 245 (S. Nadella 30(b)(1) Tr.) at 43:3–44:3; A. Blanco Opp. Decl. Ex. 281 (J. Tinter, 30(b)(6) Tr.) at 237:17-239:10 ("I and Microsoft do not believe that OpenAI has the obligation to provide us with training data."); 352:22-354:6 (listing some of the reasons OpenAI refused to provide training data). The Second Amended JDCA gave Microsoft the right to ask OpenAI questions, but OpenAI also had contractual rights to withhold information from Microsoft. Defendants further respond that the assertions in Paragraph 768 are immaterial because they do not affect any claim or defense at issue in News Plaintiffs' motion for summary judgment.
769. Microsoft could ask OpenAI what data OpenAI trained on, and OpenAI was contractually obligated to answer. Ex. 533 [Nadella Depo. Tr.] at 60:2-61:4.	769. **Disputed in Part – Improper Legal Argument; Immaterial.** Defendants do not dispute that Microsoft could ask questions to OpenAI about training data. But Defendants do dispute the degree to which OpenAI was contractually required to provide detailed information about the training data. *See* response to Paragraph No. 768. Defendants further respond that the assertions in Paragraph 769 are immaterial because they do not affect the outcome of the suit under the governing law. Microsoft's ability to ask questions about OpenAI's training is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.

770. Microsoft had the unambiguous right to download model weights (i.e., parameters) and code used by OpenAI. Ex. 514 [Tinter 30(b)(6) Depo. Tr.] at 236:6-16 ("[T]he JDCA is unambiguous that [Microsoft] have the rights to download code and it's unambiguous that [Microsoft] have the right to download model weight."); Ex. 602 [MSFT_AI_MDL_000000575] at -586; Ex. 619 [Huynh Jan. 27, 2026 30(b)(6) Depo. Tr.] at 246:4-18.	770. **Undisputed.**
771. Under the JDCA, if Microsoft had "concerns regarding the use of any particular data to train a Model," it was to notify OpenAI and OpenAI would be required to "promptly retrain that Model using solely data approved by Microsoft." Ex. 602 [MSFT_AI_MDL_000000575] at –590; *see also* Ex. 533 [Nadella Depo. Tr.] at 74:13-75:9.	771. **Disputed in Part – Immaterial.** Microsoft does not dispute that the Second Amended and Restated JDCA contains the quoted language. Microsoft disputes that the earlier JDCAs contained such language. *See* News Ex 575. Microsoft further disputes the characterization in Paragraph 771 to the extent it suggests that the provision gave Microsoft control over OpenAI's selection of training data. The provision begins: "The primary purpose of this Agreement is to provide for the Commercial Use of Licensed Products powered by Licensed IP." News Ex. 602 at MSFT_AI_MDL_000000590. And it provides that, "[s]ubject to the foregoing, OpCo and OpenAI may train all Models using data selected *solely* by OpCo and OpenAI[.]" *Id.* (emphasis added); *see* response to Paragraph No. 768. Microsoft further responds that the assertions in Paragraph 771 are immaterial because the legal conclusion regarding the scope of Microsoft's ability under the Second Amended and Restated JDCA to request OpenAI to retrain a model is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.

	OpenAI defers to Microsoft's response.
772. Microsoft's CEO Satya Nadella at deposition confirmed that, under the JDCA, "if Microsoft has any concerns regarding whether OpenAI used non public copyrighted works without permission, it can notify OpenAI and require that it promptly retrain that model." Ex. 533 [Nadella Depo. Tr.] at 74:19-25 (confirming "That is correct").	772. **Disputed in Part – Immaterial.** Microsoft does not dispute that the Second Amended and Restated JDCA contains the quoted language. But Microsoft disputes the characterization in Paragraph 772 to the extent it suggests that the provision gave Microsoft control over OpenAI's selection of training data. *See* responses to Paragraph No. 771. Microsoft further responds that the assertions in Paragraph 772 are immaterial because the legal conclusion regarding the scope of Microsoft's ability to request OpenAI to retrain a model is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment. OpenAI defers to Microsoft's response.
773. Microsoft's 30(b)(6) witness admitted that, despite concerns over copyright issues, Microsoft "instructed" its teams not to ask about OpenAI's training data. Ex. 619 [Huynh Jan. 27, 2026 30(b)(6) Depo. Tr.] at 246:4-247:6, 251:2-13.	773. **Disputed in Part – Immaterial.** Microsoft does not dispute that Jamie Huynh testified that "we always limited and instructed our teams not to ask about data." But Microsoft disputes the characterization of Paragraph 773 that Microsoft limited questions "despite concerns over copyright issues." When asked why Microsoft gave that instruction, Mr. Huynh explained that "[t]he main one was really around resources from OpenAI," because "it takes time for somebody on OpenAI to answer specific questions," and "part of the partnership was around making the partnership as efficient as possible." A.

	Blanco Opp. Decl. Ex. 214 (J. Huynh 30(b)(6) Tr. (Jan. 27, 2026)) 247:12–16. Mr. Huynh testified separately regarding: (1) awareness of copyright concerns; and (2) instructions to certain Microsoft teams not to ask about OpenAI's training data. Defendants dispute this fact as stated to the extent the SUMF's use of the word "despite" implies a causal connection between the copyright concerns and the instruction not to ask about training data. The cited testimony at News Ex. 619 at 246:4–247:6 and 251:2–13 addresses these subjects separately and does not establish that Microsoft instructed its teams not to ask about training data because of, or notwithstanding, copyright concerns. The SUMF draws an inference not supported by the cited evidence. Microsoft further responds that the assertions in Paragraph 773 are immaterial because they do not affect the outcome of the suit under the governing law. Microsoft's alleged "knowledge" about OpenAI's training data is not relevant to any asserted claim or defense at issue in News Plaintiffs' motion for summary judgment. OpenAI defers to Microsoft's response.
774. Microsoft never exercised its contractual right to require OpenAI to retrain models using data approved by Microsoft. Ex. 514 [Tinter 30(b)(6) Depo. Tr.] at 126 (Q: "Did Microsoft ever execute on its right to require OpenAI to retrain its model?" A: "No").	774. **Undisputed – Immaterial.** Undisputed, but Defendants respond that the assertions in Paragraph 774 are immaterial because they do not affect the outcome of the suit under the governing law. Whether Microsoft required OpenAI to retrain any model is not relevant to any

	asserted claim or defense at issue in News Plaintiffs' motion for summary judgment.
775. At deposition, when Microsoft's CEO Satya Nadella was asked whether he would have invoked the contractual provision requiring OpenAI to retrain its models if he "had been made aware that OpenAI had scraped and trained on information that was behind a paywall," he responded: "I would." Ex. 533 [Nadella Depo. Tr.] at 76:20-77:2.	775. **Undisputed – Immaterial; Not Supported by Admissible Evidence.** Defendants do not dispute that this question and answer appear in Mr. Nadella's deposition but Defendants respond that the hypothetical scenario described in Paragraph 775 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft further responds that Paragraph 775 is not supported by admissible evidence, as required by Local Rule 56.1. S.D.N.Y. Local Rule 56.1(d). The testimony is inadmissible because the question was an improper hypothetical, was incomplete, and called for speculation, including speculation about what was legally appropriate under the parties' agreements. In light of the objectionable question (all objections as to form were reserved), any testimony given in response is inadmissible.
776. Microsoft was aware that OpenAI used scraped news articles and books for training. *See* Ex. 619 [Huynh Jan. 27, 2026 30(b)(6) Depo. Tr.] at 169:21-24; Ex. 666 [MSFT_AI_MDL_000228652] at -652; Ex. 668 [MSFT_AI_MDL_000008148] at -148; Ex. 670 [MSFT_AI_MDL_000666181] at -181.	776. **Disputed in Part – Immaterial.** Disputed for the reasons stated in response to Paragraph 768. Microsoft further responds that the facts set forth in Paragraph 776 are immaterial to the extent Plaintiffs are suggesting that Microsoft had knowledge about what Plaintiffs claim because these facts do not affect the outcome of the suit under the governing law. Microsoft's alleged "knowledge" is not an element of any claim or defense at issue in Plaintiffs' summary judgment motion. OpenAI defers to Microsoft's response.

777. Under the Joint Development and Collaboration Agreement ("JDCA"), Microsoft had the right to terminate their agreement with OpenAI if OpenAI had breached the agreement – which would include a breach of intellectual property rights. Ex. 533 [Nadella Depo. Tr.] at 41:14-42:9; Ex. 575 [MSFT_AI_MDL_000000345] at -364.	777. **Disputed in Part – Improper Legal Argument; Not Supported by the Cited Evidence; Immaterial.** Undisputed that the JDCA provided Microsoft with a right to terminate the agreement in the event of an uncured material breach, subject to several other conditions, including the procedures set forth in Section 12(b)(ii). *See* News Ex. 575 at MSFT_AI_MDL_000000364. Neither the JDCA nor the quoted testimony support the legal argument that "a breach of intellectual property rights" allowed Microsoft to terminate the JDCA. Further, Microsoft responds that it disputes Paragraph 777 and that the Court should disregard this Paragraph because the assertion constitutes improper attorney argument. Legal conclusions regarding the scope of Microsoft's termination rights under the various versions of the JDCA are immaterial to any claim or defense at issue in News Plaintiffs' motion for summary judgment.
778. Microsoft "at any time with or without cause" retained the right to suspend or revoke OpenAI's access to the Supercomputing Technology Microsoft was providing. Ex. 672 [MSFT_AI_MDL_000003490] at -492.	778. **Disputed in Part – Not Supported by the Cited Evidence; Immaterial.** Defendants respond as follows: Undisputed that the cited amendment to the Amended JDCA provides that "Microsoft reserves the right to suspend or revoke access to the Supercomputing Tech, at any time with or without cause." However, the right to "suspend or revoke" access may also be subject to other contractual provisions—including notice requirements,

	implied covenants of good faith and fair dealing, and other terms of the parties' agreements—that the assertion in Paragraph 778 omits. *See* News Ex. 608 at MSFT_AI_MDL_000169503. Microsoft further responds that the assertions set forth in Paragraph 778 are immaterial because they do not affect the outcome of the suit under the governing law. Microsoft's ability to revoke OpenAI's access to Azure is not relevant to any asserted claim or defense at issue in News Plaintiffs' summary judgment motion.
779. Microsoft placed conditions on access to the technology they were providing OpenAI, retained the ability to terminate those services at various times and in some cases for no cause, and Microsoft never conditioned its continued provision of the supercomputing technology and infrastructure on OpenAI's compliance with copyright law. Ex. 514 [Tinter 30(b)(6) Depo. Tr.] at 274:3-275:2.	779. **Disputed in Part – Immaterial.** Microsoft does not dispute that Microsoft placed conditions on access to some of the technology it was providing OpenAI, that various agreements allowed Microsoft to terminate the agreement, sometimes without cause, and that Microsoft did not condition OpenAI's access to various technologies based on OpenAI's compliance with copyright law. But Microsoft disputes the characterization in Paragraph 779 of the scope of those conditions and termination rights afforded to Microsoft, including as being inclusive of IP violations. Mr. Tinter testified that the conditions served to "protect some of the technology that we brought to the collaboration," and that while [redacted] A. Blanco Opp. Decl. Ex. 281 (J. Tinter 30(b)(6) Tr.) 274:19–22).

	Microsoft further responds that the assertions set forth in Paragraph 779 are immaterial because they do not affect the outcome of the suit under the governing law. Microsoft's ability to revoke OpenAI's access to Azure and supercomputers is not relevant to any asserted claim or defense at issue in News Plaintiffs' summary judgment motion. OpenAI defers to Microsoft's response.
780. OpenAI was contractually required to "exclusively use Microsoft Azure and other relevant Microsoft supercomputing services for all AI training inferenced deployment and commercial workloads." Ex. 514 [Tinter 30(b)(6) Depo. Tr.] at 337:10-21; Ex. 674 [OPCO_MDL_PD_000000432] at -432.	780. **Undisputed.** Defendants do not dispute that OpenAI was to use Microsoft's Azure Services for AI training, inference, deployment, and commercial workloads throughout the term of the Amended JDCA. However, Paragraph 780 cites the wrong document for this fact. News Ex. 674 is a draft term sheet, not an executed JDCA. Although Mr. Tinter testified that he believed the provision was included in the final terms, the document he reviewed was likewise a draft. Ex. 514 at 337:18–21.

III. Defendants' Use and Copying of News Content

781. Defendants used and copied news content at different stages, including: (1) data acquisition and processing, (2) training, including pre-training, mid-training, and post-training, (3) retrieval augmented generation, (4) outputs, and (5) distributing data between each other.	781. **Disputed - Not Supported by Cited Evidence.** Defendants respond as follows: The facts set forth in Paragraph 781 are disputed because they are not supported by any cited evidence. Further disputed to the extent Plaintiffs' characterization conflates distinct technical processes and assumes, without evidentiary

	support, that each involves copying of Plaintiffs' Asserted Works.

A. Acquisition of News Content

782. One cannot train on data without first acquiring it in digital form—typically by downloading it from the Internet—which requires making a copy of it. Ex. 676 [Dr. Callison-Burch Rebuttal Rep.] at ¶¶ 59-64.	782. **Undisputed.** Microsoft defers to OpenAI.
783. Defendants acquired much of the text used to train the models at issue either by directly scraping content from the internet using a scraper bot or acquiring web content that is scraped by a third-party like Common Crawl. Ex. 678 [OPCO_MDL_004806204] at –205, –213.	783. **Disputed in Part — Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that [redacted], as stated in the cited document. Disputed to the extent that "much" of the entire set of training data was obtained by scraping. Microsoft disputes Paragraph 783 to the extent it asserts that Microsoft "acquired much of the text used to train the models at issue." Microsoft's models are not at issue in this case, nor did Microsoft train the models at issue in this case. The cited evidence does not support that assertion as to Microsoft. News Ex. 678 is an OpenAI document describing the dataset OpenAI used to train GPT-4. It does not mention Microsoft acquiring any training text. News Ex. 678 at OPCO_MDL_004806204–05.

784. A bot (short for robot) is a software program that performs tasks over a network in an automated fashion. Ex. 410 [Feamster Opening Report] ¶ 110.	784. **Undisputed.** Microsoft defers to OpenAI.
785. A scraper is a bot that extracts data from a webpage. Ex. 410 [Feamster Opening Report] ¶ 109.	785. **Undisputed.** Microsoft defers to OpenAI.
786. On several occasions, OpenAI operated its own web crawlers and web scrapers to directly scrape content from the internet. Ex. 627 [Ian Sohl, March 31, 2025 30(b)(1) Deposition Tr.] at 37:17-38:19.	786. **Undisputed.** Microsoft defers to OpenAI.
787. Scraping News Plaintiffs' content, and acquiring News Plaintiffs' content scraped by others, requires copying that content. Ex. 676 [Callison-Burch Rebuttal Report] ¶ 59.	787. **Disputed in Part — Not Supported by Cited Evidence; Improper Legal Argument.** Web scraping, data processing, filtering, and tokenization are general technical processes that may be used to acquire and prepare data for model training. OpenAI disputes any characterization of those processes as specific to News Plaintiffs' content as such characterizations are not supported by the cited report by Callison-Burch. Callison-Burch's description of those processes is not an admission that OpenAI scraped News Plaintiffs' content or acquired scraped copies of that content. OpenAI further disputes any suggestion that each such process necessarily involves copying within the meaning of the Copyright Act. News Ex. 676 (Callison-Burch Rebuttal Report) ¶ 59. Microsoft defers to OpenAI.

788. OpenAI prioritized acquiring news content for use in training and developing its large language models. Ex. 688 [OPCO_MDL_000113921] at -25 (explaining how OpenAI president Greg Brockman "scraped" websites including [redacted]); Ex. 689 [OPCO_MDL_000156469] at -72 (Mr. Brockman in January 2022 describing "my targeted scrape of high quality sites"); Ex. 551 [OPCO_MDL_000481178] at -78 (February 2022 conversation noting that "greg [Brockman] has prepped some high quality, small datasets" including [redacted]); Ex. 690 [OPCO_MDL_000113767] at -67 (Mr. Brockman referring to "my scraped data ([redacted]); Ex. 692 [OPCO_MDL_001820445] at -46 (Mr. Brockman stating: "I think I can get [redacted] my sitemap scrape without breaking a sweat should be a much denser scrape than [redacted]); Ex. 694 [OPCO_MDL_001841113] at -13 (Mr. Brockman's colleague asking about his "bunch of higher quality/smaller web scrape datasets [redacted]); Ex. 696 [OPCO_MDL_000156332] at -32 (Mr. Brockman writing in January 2022 that "i completed my sitemap for [redacted] and am scraping their site now").	788. **Disputed — Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that the cited evidence establishes that OpenAI generally "prioritized" acquiring news content for training and developing its large language models. The cited documents concern Brockman's efforts to obtain data from a variety of websites, including news publishers, Medium, and Substack, based on their perceived quality, but do not establish that news content was prioritized over other categories of training data. Indeed, News Plaintiffs acknowledge in Paragraphs 867–868 that news content is interchangeable with other types of data for purposes of training LLMs. Microsoft defers to OpenAI.
789. OpenAI scraped websites, including news websites, to obtain content to use for training and developing its large language models. See evidence cited in prior Fact.	789. **Undisputed.** OpenAI further responds to note, however, that the "evidence cited in prior Fact," does not itself constitute evidence; to the extent the previous fact is disputed, so too is this summary assertion.

	Microsoft defers to OpenAI.
790. OpenAI devised and implemented strategies to prevent websites from knowing that OpenAI was scraping their content. Ex. 698 [OPCO_MDL_004639273] at -73 (In January 2022, OpenAI's President Greg Brockman and then-colleague Alec Radford discussing strategies for avoiding detection by sites while scraping, with Mr. Radford advising to "estimate their traffic and make sure your never more than a like 1% bump or something"); Ex. 700 [OPCO_MDL_004836003] (OpenAI employee Vineet Kosaraju discussing "the sites we want to scrape" and his "rough sense of [their] anti-bot measures" and explaining how those measures "should mostly be addressable"); Ex. 702 [OPCO_MDL_001940220] at -20 (In January 2023, OpenAI's Chief Research Officer Bob McGrew describing how "our web scrape is going slower than we expected because we are seeing our crawlers blocked from many large content providers" while assuring colleagues that "[w]e are working around this by slowing down our scrape"); Ex. 704 [OPCO_MDL_001937738] at -46_0005 (October 2019 document stating that "Up till now basically all of our datasets have come from scraping the internet" and that "we're kinda flagrantly going around rate limits and generally very a bit aggressive when it comes to scraping", and questioning whether "we can commercialize products obtained via scraping").	790. **Disputed — Not Supported by Cited Evidence.** OpenAI responds as follows: The cited evidence does not establish that OpenAI devised and implemented strategies to prevent websites from knowing that OpenAI was scraping their content. Rather, the cited documents concern discussions of crawl rates, website traffic, anti-bot measures, and technical challenges encountered during web scraping. A. Blanco Opp. Decl. Ex. 176 (G. Brockman 30(b)(1) Tr.) 335:11-15. Microsoft defers to OpenAI.

791. Some OpenAI employees devised and implemented strategies to circumvent paywalls when scraping websites. Ex. 706 [OPCO_MDL_001940102] at -03 (Nick Ryder informing Mr. Brockman about "a hack to get around nytimes paywall" to help with Mr. Brockman's efforts to "scrap[e]" The Times's site, and Mr. Brockman responding "ah nice").	791. **Disputed in Part – Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that certain OpenAI employees discussed paywalls related to websites and scraping. However, disputed that the cited evidence establishes that OpenAI employees devised and implemented strategies to circumvent paywalls when scraping websites. The cited exchange reflects a discussion of a potential method for accessing content behind The Times' paywall, but does not establish that the method was implemented, used to acquire training data, or adopted as part of OpenAI's data acquisition practices. Microsoft defers to OpenAI.
792. OpenAI disregarded paywalls when scraping websites. Ex. 708 [OPCO_MDL_001686009] at -09 (OpenAI employee in August 2023 answering "yes" when asked "do we also scrape paywalled sites"); Ex. 414 [Sohl Dec. 9. 2025 30(b)(6) Tr.] at 368:2-11 ("Q. Does OpenAI make any effort to detect paywall content in its training datasets? A. Not to my knowledge." "Q. Does OpenAI make any effort to remove paywall content from its training datasets? A. Not to my knowledge.").	792. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 792 are disputed because they are not supported by the record. Specifically, the cited evidence does not support the assertion that OpenAI "disregarded" paywalls when scraping websites. The cited Slack message and deposition testimony address whether OpenAI scraped paywalled sites and whether OpenAI made efforts to detect or remove paywall content, but do not establish that OpenAI affirmatively "disregarded" paywalls. To the contrary, Greg Brockman testified that he is not aware of anybody at OpenAI who made any attempts to "work around New York Times' paywalls." A. Blanco Opp. Decl. Ex. 176 (G. Brockman 30(b)(1) Tr.) 348:21-349:3.

	Microsoft defers to OpenAI.
793. At his deposition, Mr. Nadella testified that "anything that is paywalled should be licensed by anyone who wants to use it for whether grounding or training." Ex. 533 [Nadella Depo. Tr.] at 96:2-10; 102:11-15 ("Q: Would you agree that you should not train on New York Times paywalled content without paying for the right to do so? A: I agree with that.").	793. **Undisputed – Incomplete; Immaterial.** Microsoft does not dispute that Exhibit 533 contains the quoted language. The questions were objectionable because they called for a legal conclusion, and the answers are therefore inadmissible. Further, pursuant to Fed. R. Evid. 106, the first quoted answer was in response to a question limited to "content that is paywalled and unlicensed to start." A. Blanco Opp. Decl. Ex. 245 (S. Nadella 30(b)(1) Tr.) 96:2–10. Mr. Nadella also clearly limited his testimony to his personal opinion by disclaiming any legal meaning for his views: "I don't have the legal definitions of all of this" and "not a legal term, but how I think about [it]." *Id.* 102:20–21, 103:1 Microsoft also responds that the facts stated in Paragraph 793 are immaterial because these facts do not affect the outcome of any claim or defense at issue in News Plaintiffs' motion for summary judgment. Specifically, personal opinions are not relevant to the four factor of fair use. OpenAI defers to Microsoft's response.
794. At this deposition, Mr. Nadella confirmed that "something that you shouldn't do is to train on any data that's not public," including "data that is behind a paywall" as well as "accessing data in violation of a term of service." Ex. 533 [Nadella Depo. Tr.] at 84:12-85:16.	794. **Disputed in Part – Immaterial; Not Supported by Cited Evidence; Incomplete.** Microsoft does not dispute that Mr. Nadella testified that "from a training perspective, what I would consider something that you shouldn't do is to train on any data

	that's not public," and agreed that "non public" includes "data that is behind a paywall." A. Blanco Opp. Decl. Ex. 245 (S. Nadella 30(b)(1) Tr.) 84:12–19. But Microsoft disputes that Mr. Nadella testified that training on data "in violation of a term of service" is something "you shouldn't do" because the cited testimony is taken out of order, and the question was objectionable because it clearly called for an improper legal conclusion. The testimony is inadmissible and thereby does not support the stated fact. Further, pursuant to Fed. R. Evid. 106, when asked whether "non public" data includes data that a website's terms of service do not allow to be scraped, Mr. Nadella testified: "The scraping is a different problem which is scraping is not a training issue, it's more an abuse of- because you blocked them, right? . . . it's a nuisance that your computational resources are being used by somebody who is scraping who is not a user." *Id.* 84:20–85:6. Mr. Nadella's agreement that the "problems can include accessing data in violation of a term of service" followed *that* answer. *Id.* 85:7–11. Microsoft also responds that the facts set forth in Paragraph 794 are immaterial for the reasons stated in response to Paragraph 793. OpenAI defers to Microsoft's response.
795. At his deposition, Mr. Nadella agreed that AI companies should not "access[] data in violation of a term of service." Ex. 533 [Nadella Depo. Tr.] at 85:7-11.	795. **Disputed in Part – Immaterial; Not Supported by Cited Evidence.** Microsoft does not dispute that Mr. Nadella agreed that "the problems can include accessing data in violation of a term of service." A.

	Blanco Opp. Decl. Ex. 245 (S. Nadella 30(b)(1) Tr.) 85:7–11. But Microsoft disputes that Mr. Nadella "agreed that AI companies should not" do so because the cited testimony does not support that characterization. Pursuant to Fed. R. Evid. 106, the question did not refer to AI companies, and it followed Mr. Nadella's testimony that "scraping is not a training issue, it's more an abuse of- because you blocked them," and "a nuisance that your computational resources are being used by somebody who is scraping who is not a user." *Id.* at 84:20–85:6; *see also id.* 83:19–84:3 (explaining that scraping in violation of terms of service raises computational-cost concerns and is "a different issue"). Additionally, these questions improperly called for legal conclusions and any responses are therefore inadmissible. Microsoft also responds that the facts set forth in Paragraph 795 are immaterial for the reasons stated in response to Paragraph 793. And, Mr. Nadella's testimony concerned the computational burden of unwanted scraping which is immaterial to the issues before the Court. *Id.* 83:19–84:3, 84:24–85:6. OpenAI defers to Microsoft's response.
796. Mr. Nadella confirmed that if he "had been made aware that OpenAI had scraped and trained on information that was behind a paywall," he would have "invoked [Microsoft's right to] require[] OpenAI to retrain its models." Ex. 533 [Nadella Tr.] at 76:20-77:2 (Q: "If you, Mr.	796. **Undisputed – Immaterial; Not Supported by Admissible Evidence.** Microsoft does not dispute that this question and answer appear in Mr. Nadella's deposition but Microsoft responds that the hypothetical scenario described in Paragraph 796 is

Nadella, had been made aware that OpenAI had scraped and trained on information that was behind a paywall, would you have invoked this provision that required OpenAI to retrain its models? A: I would.").	immaterial because it does not affect the outcome of the suit under the governing law. Microsoft further responds that Paragraph 796 is not supported by admissible evidence, as required by Local Rule 56.1. S.D.N.Y. Local Rule 56.1(d). The testimony is inadmissible because the question was an improper hypothetical, was incomplete, and called for speculation, including speculation about what was legally appropriate under the parties' agreements. In light of the objectionable question (all objections as to form were reserved), any testimony given in response is inadmissible. OpenAI defers to Microsoft's response.
797. In an October 27, 2023 slack message, Mira Murati, OpenAI's CTO at the time, stated that "if a site prohibits OAI from using its content and OAI uses it anyway, it certainly can create risks, right?" Ex. 710 [OPCO_MDL_000000804] at -805.	797. **Disputed in Part**. Defendants respond as follows: Undisputed that Ex. 710 contains the quoted language in a Slack message sent by Murati. However, disputed to the extent Plaintiffs characterize the quoted language as Murati's own statement. The quoted language originated from Microsoft employee Mikhail Parakhin and was relayed by Murati, rather than independently authored by her. Further disputed that the message was originally sent on October 27, 2023, as News Ex. 710 reflects a September 27, 2023 timestamp for the original message. Microsoft further responds as follows: Microsoft does not dispute that the quoted language originated in a message from Mr. Parakhin, which Ms. Murati later relayed. Microsoft disputes Plaintiffs' characterization to the extent

	it omits the rest of the statement. Pursuant to Fed. R. Evid. 106, the full message states that "it really is OAI risks we are talking about, not Microsoft's," and that the "OAI crawler" is "the way to discover whether the websites prohibit OAI or not." News Ex. 710 at -OPCO_MDL_000000805. OpenAI further responds as follows: OpenAI disputes that the referenced message establishes that OpenAI's bots accessed sites that prohibited them. Specifically, the referenced message does not support any assertion that OpenAI's bots ignored robots.txt directives or failed to follow implemented guardrails. Dr. Feamster explained that OpenAI publishes the user agent identifiers ChatGPT-User and OAI-SearchBot, which allows website owners to manage those bots' access through their robots.txt files. *See* N. Feamster Opp. Decl. ¶¶ 48-49. Dr. Feamster also analyzed the designed behavior of GPTBot by reviewing its source code and the practical behavior of GPTBot by analyzing server logs, and concluded that GPTBot's practice was consistent with the Robots Exclusion Protocol. *See id.* ¶¶ 41-43. In fact, Plaintiffs' own class expert witnesses Shan and Zhao confirmed that GPTBot respects robots.txt in practice. Shan & Zhao, Somesite I Used to Crawl, https://arxiv.org/html/2411.15091v1, § 5.2.1. Further, to the extent Plaintiffs rely on the referenced message to characterize OpenAI's RAG processes, OpenAI responds to note that Browse was designed such that its

	bots do not fetch content from websites whose robots.txt files indicate a contrary preference. *See* ECF 1756 (N. Feamster Decl.) ¶¶ 50-51. OpenAI also performs a robots.txt check before Browse sends a request to a webpage for direct retrieval. *See id.* ¶¶ 16, 51. OpenAI did not visit News Plaintiffs' sites against their robots.txt preferences. OpenAI incorporates its response to News Fact 1559 by reference.
798. OpenAI represented to The Times in April 2023 that OpenAI "has not accessed, NYT platforms (including its websites and mobile applications) for AI training or any related purpose." Ex. 139 [OPCO_MDL_004621506] at -07.	798. **Disputed in Part.** OpenAI responds as follows: Undisputed that Ex. 139 contains the quoted language. However, disputed to the extent Plaintiffs suggest that OpenAI represented its training datasets did not contain any New York Times content. The quoted statement was made in response to the NYT's specific question on whether OpenAI was accessing its platforms, including websites and mobile applications, through crawling, scraping, or otherwise. However, in response to a separate question about whether OpenAI was using NYT content "from any source" for AI training or related purposes, OpenAI explained that its "research and training data includes publicly available sources like Common Crawl and Wikipedia, which may include publicly available portions of NYT content," and noted that OpenAI did not seek out such content for research and training purposes. News Ex. 139 at OPCO_MDL_004621507. Microsoft defers to OpenAI.

799. Prior to April 2023, OpenAI had scraped The Times's site to obtain content for use in training and/or developing its large language models. Ex. 414 [Sohl Dec. 9. 2025 30(b)(6) Tr.] at 290:11-16, 332:6-17 (OpenAI corporate representative confirming at deposition that "OpenAI directly scraped The Times's website", and when confronted with OpenAI's April 2023 email to The Times, admitting "I am clear that OpenAI did perform a scrape of content from New York Times before this date").	799. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI directly scraped The Times' website before April 2023. However, disputed that the cited evidence establishes that OpenAI did so to obtain content for training or developing the at-issue LLMs. Pursuant to Fed. R. Evid. 106, Sohl testified that, to his understanding, "none of the production models at question were trained using the data that was referenced in this particular scrape," and that he was not aware of any other New York Times-specific datasets used to train those production models. A. Blanco Opp. Decl. Ex. 277 (I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) 296:24–297:16. Further, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
800. OpenAI lied in April 2023 when it told The Times that OpenAI "has not accessed, NYT platforms (including its websites and mobile applications) for AI training or any related purpose." Ex. 139 [OPCO_MDL_004621506] at -07; Ex. 414 [Sohl Dec. 9. 2025 30(b)(6) Tr.] at 290:11-16, 332:6-17 (OpenAI employee confirming at deposition that "OpenAI directly scraped The Times's website", and when confronted with OpenAI's April 2023 email to The Times, admitting "I am clear that OpenAI did perform a scrape of content from New York Times before this date").	800. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The cited evidence does not establish that OpenAI knowingly made a false representation to The Times in April 2023. Although Mr. Sohl acknowledged that OpenAI had previously scraped content from The Times' website, his testimony does not establish that the scraping was conducted "for AI training or any related purpose." Indeed, pursuant to Fed. R. Evid. 106, Mr. Sohl testified that, to his understanding, "none of the production models at question were trained using the data that was referenced in this particular scrape," and that he was not aware of any other New York

	Times-specific datasets used to train those production models. A. Blanco Opp. Decl. Ex. 277 (I. Sohl 30(b)(6) Tr. (Dec. 9, 2025)) 296:24–297:16. The facts set forth in Paragraph 800 are disputed because the cited evidence is incorrect. Microsoft defers to OpenAI.

1. WebText and WebText2

801. WebText is a scrape of internet URLs that OpenAI created in 2019 as part of the development of GPT-2. Ex. 456 [Radford Feb. 24, 2026 Depo. Tr.] at 238:19-24.	801. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that WebText is a scrape of Internet URLs that OpenAI performed as part of the development of GPT-2. Disputed that the scrape occurred in 2019 as contrary to the record. Pursuant to Fed. R. Evid. 106, *see* A. Blanco Opp. Decl. Ex. 261 (A. Radford 30(b)(1) Tr.) 410:5-7. Further, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
802. OpenAI compiled the WebText and WebText2 datasets by scraping all outbound links from Reddit that received at least 3 karma (i.e., positive upvotes). Ex. 732 [OPCO_MDL_000200524] at -26; *see also* Ex. 414 [Dec. 9, 2025 Depo. Tr. of Ian Sohl] at 297:21-25; Ex. 731 [OPCO_MDL_001687062] at -66.	802. **Undisputed.** Microsoft defers to OpenAI.

803. WebText "emphasizes document quality" by prioritizing web pages "curated/filtered by humans." Ex. 644 [Language Models Are Unsupervised Learners Paper, https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf] at 3.	803. **Disputed in Part.** OpenAI responds as follows: Undisputed that the authors of News Ex. 732 described how they "created a new web scrape [WebText] which emphasizes document quality" and further described that, "[t]o do this [they] only scraped web pages which have been curated/filtered by humans" but disputed to the extent the statement suggests documents curated or filtered by humans are necessarily high quality for AI development because the evidence does not support that assertion. Microsoft defers to OpenAI.
804. News articles were the most prevalent type of content in WebText. Ex. 733 [OPCO_MDL_004635615].	804. **Disputed.** OpenAI responds as follows: Disputed that the cited evidence establishes that news articles were the most prevalent type of content in WebText. The cited document does not quantify the proportion of WebText consisting of news articles or establish how news articles were categorized relative to other types of content. Indeed, pursuant to Fed. R. Evid. 106, Radford testified [redacted]. A. Blanco Opp. Decl. Ex. 261 (A. Radford 30(b)(1) Tr.) 251:22–253:25. Microsoft defers to OpenAI.

805. Webtext2 is an expanded version of WebText compiled by scraping more Reddit outlinks. Ex. 456 [Radford Feb. 24, 2026 Depo. Tr.] at 247:22-248:19.	805. **Disputed in Part.** OpenAI responds as follows: Undisputed that Mr. Radford testified that the creation of WebText2 may have involved scraping additional links. However, disputed that the cited testimony establishes that WebText2 was an expanded version of WebText compiled by scraping additional Reddit outlinks. Pursuant to Fed. R. Evid. 106, Mr. Radford [redacted] A. Blanco Opp. Decl. Ex. 261 (A. Radford 30(b)(1) Tr.) 247:2–7, 248:20–23. Microsoft defers to OpenAI.
806. OpenAI scraped a total of 168,091 Ziff Davis webpages in webtext. Ex. 655 [number of webpages scraped per domain downloaded from https://github.com/openai/gpt-2/blob/master/domains.txt]; *see also* Dkt. 1676 ¶ 136.	806. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI employees published a list characterized as identifying the top 1,000 domains represented in WebText and their respective frequencies, which includes entries for certain websites associated with Ziff Davis. Disputed that the cited evidence establishes that OpenAI scraped a total of 168,091 Ziff Davis webpages in WebText. The cited list identifies domain-level frequencies but does not establish that each entry represents a unique webpage scraped by OpenAI, that the identified domains were all owned by Ziff Davis during the relevant period, or that the sum of those frequencies represents the total number of Ziff Davis webpages scraped and included in WebText. Further, the

	disputed proposition is not supported by the cited evidence. Microsoft responds that these facts do not pertain to claims asserted against Microsoft.
807. The breakdown of Ziff Davis domains scraped in webtext is as follows: Domain — Webpages news.com — 27112 cnet.com — 29710 mashable.com — 20584 ign.com — 40071 humblebundle.com — 15132 lifehacker.com — 10032 zdnet.com — 12171 vg247.com — 5905 rockpapershotgun.com — 7374 **Total — 168091**	807. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI employees published a list characterized as identifying the top 1,000 domains represented in WebText and their respective frequencies, which includes entries for certain websites associated with Ziff Davis. Disputed that the cited evidence establishes that OpenAI scraped a total of 168,091 Ziff Davis webpages in WebText. The cited list identifies domain-level frequencies but does not establish that each entry represents a unique webpage scraped by OpenAI, that the identified domains were all owned by Ziff Davis during the relevant period, or that the sum of those frequencies represents the total number of Ziff Davis webpages scraped and included in WebText. Further, the disputed proposition is not supported by the cited evidence. Microsoft responds that these facts do not pertain to claims asserted against Microsoft.

Ex. 655 [number of webpages scraped per domain downloaded from https://github.com/openai/gpt-2/blob/master/domains.txt].	
808. OpenAI's employees and internal documents recognized that news articles "were prevalent" in Webtext and WebText2. Ex. 529 [Chess Tr.] at 161:25-162:5 ("They were prevalent."), *id.* 162:6-16 ("I remember seeing New York Times come up a bunch…."); *see also* Ex. 733 [OPCO_MDL_004635615].	808. **Disputed in Part.** OpenAI responds as follows: Undisputed that News Ex. 529 contains the quoted testimony. Disputed that the cited evidence establishes the composition of Webtext and Webtext2. Moreover, pursuant to Fed. R. Evid.106, Mr. Radford testified [redacted] A. Blanco Opp. Decl. Ex. 261 (A. Radford 30(b)(1) Tr.) 251:22–253:25. Microsoft defers to OpenAI.
809. OpenAI's WebText2 scraper scraped a webpage's entire HTML. Wenger Opening Rep. at 36.	809. **Disputed in Part - Expert Opinion Subject to Rule 702.** OpenAI responds as follows: Disputed that the cited evidence establishes that OpenAI's WebText2 scraper necessarily scraped a webpage's entire HTML. Dr. Wenger states [redacted]" ECF 1858-1 (E. Wenger Decl. Ex. A) at 36 (emphasis added). Plaintiffs' characterization omits the qualification "[redacted]" and improperly suggests that the scraper invariably collected the entirety of each webpage's HTML, which Dr. Wenger's opinion does not establish. Further, the evidence cited in support of the disputed proposition is currently subject to a motion to exclude expert testimony pursuant to Fed. R.

	Evid. 702. Defendants have moved to exclude certain opinions of Dr. Emily Wenger (text-matching, "regurgitation," CMI-removal, and Copilot log extrapolation opinions). *See* ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 19-22. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
810. WebText contains copies of News Plaintiffs' Asserted Works. *See infra* Section re: "Defendants' Infringement of News Plaintiffs' Asserted Works Ripe for Summary Judgment"	810. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Plaintiffs cite no specific record evidence supporting Fact No. 810, as required by Fed. R. Civ. P. 56(c)(1). Plaintiffs' general cross-reference to a separate section of their Rule 56.1 statement, without identifying particular evidence supporting this assertion, is insufficient to establish the asserted fact. The fact set forth in Paragraph 810 is disputed because the cited evidence is incorrect. Microsoft defers to OpenAI.
811. WebText2 contains copies of News Plaintiffs' Asserted Works. *See infra* Section re: "Defendants' Infringement of News Plaintiffs' Asserted Works Ripe for Summary Judgment"	811. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Plaintiffs cite no specific record evidence supporting Fact No. 811, as required by Fed. R. Civ. P. 56(c)(1). Plaintiffs' general cross-reference to a separate section of their Rule 56.1 statement, without identifying particular evidence supporting this assertion, is insufficient to establish the

	asserted fact. The fact set forth in Paragraph 811 is disputed because the cited evidence is incorrect Microsoft defers to OpenAI.

2. Common Crawl

812. The largest single source of data in OpenAI's web data for the models at issue is Common Crawl. Ex. 678 [OPCO_MDL_004806204] at –205 ("[redacted] The two largest sources of data were Common Crawl [redacted] *Id.*	812. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that the cited support is accurate as to GPT-4. News Ex. 678 at OPCO_MDL_004806205. However, disputed given that this assertion as written is (incorrectly) generally applicable to all models, because that assertion is not supported by the cited evidence. Specifically, the statement is inaccurate for any models that follow GPT-4 or that were trained using the GPTBot crawls. Microsoft defers to OpenAI.
813. OpenAI downloaded the Common Crawl dataset multiple times over a period of years. Ex. 414 [Ian Sohl, December 9, 2025, 30(b)(6) Depo. Tr.] at 260:25-261:10.	813. **Undisputed.** Microsoft defers to OpenAI.
814. Common Crawl scrapes webpages and releases raw versions of those scrapes on its website. Wenger Opening Rep. at 14-15.	814. **Undisputed.** Microsoft defers to OpenAI.
815. Common Crawl has a Terms of Use, which provides: "You agree that your use of the Service must comply with all applicable local, state, national, and	815. **Undisputed.** Microsoft defers to OpenAI.

international laws, and that you will not use the Service for any illegal purpose. You also acknowledge and agree that all information, data, text, scripts, web pages, web sites, software, html page links, open data APIs, metadata or other materials (collectively, the 'Crawled Content') may be subject to separate terms of use or terms of service from the owners of such Crawled Content. You acknowledge and agree to follow any technological implementations restricting how you may used Crawled Content, including without limitation, the use of robot.txt files and NOFOLLOW meta-tags." Ex. 746 [Sohl Ex. 406] at 1.	
816. Common Crawl's web scraper scrapes and makes publicly available the entire HTML of the scraped webpages. Wenger Opening Rep. at 30-35.	816. **Undisputed.** Microsoft defers to OpenAI.
817. Plaintiffs' Expert Dr. Goldstein reviewed Common Crawl's scraper code and identified "no evidence in the code that it respects paywalled content." Goldstein Rep. ¶ 532.	817. **Disputed in Part - Expert Opinion Subject to Rule 702.** OpenAI responds as follows: Undisputed that paragraph 532 of Dr. Goldstein Opening Report contains the quoted statement. Disputed because, as OpenAI expert Dr. Feamster explains, Dr. Goldstein's characterization "misunderstands how content access restrictions work on the Internet," because there is no industry standard requiring web crawlers to detect and respect client-side subscription enforcement controls. News Ex. 423 (N. Feamster Rebuttal Rep.) ¶ 144. Feamster further explains that imposing such a requirement would be technically impracticable and is not observed by major web crawlers. *Id.* Further, the evidence cited in support of the disputed proposition is currently

	subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Defendants have moved to exclude certain opinions of Dr. Thomas Goldstein ("copies," CMI-removal, Copilot log, and Microsoft-role opinions). *See* ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 9-19. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
818. The scraped content that OpenAI collected from Common Crawl was saved in a WARC file format which contains full HTML documents that mirror the raw form of a webpage. Ex. 521[Yuan Depo. Tr.] at 21:22-22:18.	818. **Undisputed.** Microsoft defers to OpenAI.
819. When OpenAI obtained scraped content from Common Crawl it downloaded the whole WARC file. Ex. 521 [Yuan Depo. Tr.] at 22:6-9.	819. **Undisputed.** Microsoft defers to OpenAI.
820. OpenAI stores data downloaded from Common Crawl in its Azure storage account. Ex. 521 [Yuan Depo. Tr.] at 21:22-22:16; Ex. 631 [Frederiksen-Cross Depo. Tr.] at 129:4-11, 131:10-20.	820. **Undisputed.**
821. The same webpage can appear in multiple datasets made available by Common Crawl. Wenger Opening Rep. at 33.	821. **Undisputed.** Microsoft defers to OpenAI.

822. OpenAI's plan was to retain the raw Common Crawl WARC files "forever" so that OpenAI could recreate the datasets. Ex. 521 [Yuan Depo. Tr.] at 26:16-27:8; Ex. 735 [OPCO_MDL_000205483] at -486.	822. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI retained raw Common Crawl WARC files and that such files could be used to recreate datasets. However, disputed that it was "OpenAI's plan" to retain these files "forever." The cited evidence does not support this assertion. Specifically, the cited evidence reflects one engineer's subjective statement in 2021 and corresponding 30(b)(1) testimony, which does not support an unbounded statement of OpenAI's corporate intent to retain files "forever." *See* A. Blanco Opp. Decl. Ex. 291 (Q. Yuan 30(b)(1) Tr.) 26:16-27:8; News Ex. 735 at OPCO_MDL_000205486. Microsoft defers to OpenAI.
823. For Common Crawl-sourced datasets, OpenAI first acquired an initial dataset containing raw HTML content scraped from the Internet by Common Crawl and then performed processing on that dataset to create a new dataset for training. Goldstein Opening Report ¶¶ 137-139, 520; Wenger Opening Rep. at 28, 48; Delorey Opening Report ¶ 229; Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17 ("The first step in the data processing workflow is to extract usable language data from raw HTML data that OpenAI has acquired."); Ex. 750 [OpenAI's Supplemental Response to Daily News Interrogatory No. 10] (same); Ex. 752 [OpenAI's Supplemental Response to ZD Interrogatory No. 20	823. **Undisputed.** Microsoft defers to OpenAI, but also responds that facts related solely to Ziff Davis do not pertain to claims asserted against Microsoft.

(2/27/26)] at 14 (same); Ex. 754 [January 27, 2026 Email from Tom Gorman to Stephen Match] (identifying specific date ranges from Common Crawl for different datasets that OpenAI obtained).	
824. OpenAI's Common Crawl datasets contain copies of News Plaintiffs' Asserted Works. *See infra* Section re: "Defendants' Infringement of News Plaintiffs' Asserted Works Ripe for Summary Judgment"	824. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Plaintiffs cite no specific record evidence supported Fact No. 824, as required by Fed. R. Civ. P. 56(c)(1). Plaintiffs' general cross-reference to a separate section of their Rule 56.1 statement, without identifying particular evidence supporting this assertion, is insufficient to establish the asserted fact. Further, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
825. OpenAI and Microsoft have provided funding to Common Crawl. Ex. 737 [OPCO_MDL_004424027] (September 2023 grant in which OpenAI agreed to give Common Crawl $750,000); Ex. 414 [Sohl Dec. 9. 2025 (30)(6) Tr.] at 278:2-3; Ex. 1259 [OPCO_MDL_004602292] (July through September 2023 emails between OpenAI and Common Crawl discussing Common Crawl's "request to OpenAI for support"); Ex. 1135 [Common Crawl 2013 Form 990-PF] (Common Crawl tax filing showing $120,000 contribution from Microsoft in 2013); Ex. 1136 [Common Crawl 2014 Form 990-PF] (Common Crawl tax filing showing $35,000 contribution from Microsoft in 2014); Ex. 1267 [Common Crawl 2024	825. **Disputed in Part – Not Supported by Cited Evidence; Not Supported by Admissible Evidence.** OpenAI responds as follows: Undisputed that OpenAI provided several grants to Common Crawl. However, disputed to the extent that Plaintiffs imply that these grants were payment for receiving data from Common Crawl or payments for future services from Common Crawl. The funding was a charitable grant to a 501(c)(3) nonprofit organization, not as a license fee or other consideration for data rights. News Ex. 737 at OPCO_MDL_004424027. Further, OpenAI disputes any implication that this funding reflects any understanding that OpenAI was required to pay, or should have paid, for

Form 990-PF] (Common Crawl tax filing showing $20,000 contribution from Microsoft in 2024).	the right to use Common Crawl's data which it makes available for free. OpenAI further disputes any implication that OpenAI paid Common Crawl to expand the depth or breadth of its crawls in order to increase the volume of available data for training. OpenAI's corporate designee testified to his understanding that, at the time the donations were made, an increase in Common Crawl data "would not have significantly influenced the total size of the datasets used to train our models." A. Blanco Opp. Decl. Ex. 277 (I. Sohl (30)(6) Tr. (Dec. 9. 2025)) 283:23-284:2. Microsoft defers to OpenAI regarding the grants that OpenAI provided to Common Crawl. Microsoft further responds as follows: The statement that "Microsoft [has] provided funding to Common Crawl is not supported by admissible evidence. Exhibits 1135, 1136, and 1267 are unauthenticated and inadmissible hearsay. Further, Microsoft disputes that the cited evidence suppors any implication that any such "funding" from Microsoft to Common Crawl, a 501(c)(3) nonprofit organization, constituted payments for receiving data from Common Crawl or payments for future services from Common Crawl.

3. Project Mango Crawl

826. As explained above, under Project Mango, OpenAI and Microsoft collaborated to scrape web content. Ex. 617 [Canel 30(b)(6) Deposition Tr.] at 97:11-98:3; Ex.	826. **Disputed in Part – Not Supported by Cited Evidence.** While Defendants do not dispute that Microsoft and OpenAI collaborated to collect web content under

613[Gaggara 1/27/26 Tr.] at 140:10-13; Ex. 619 [Huynh 30(b)(6) Deposition Tr.] at 65; Ex. 632 [MSFT_AI_MDL_000458392]; *See* Section II.C.	Project Mango, Defendants dispute that Microsoft and OpenAI collaborated to "scrape" web content. Under the parties' Web Crawling Services Agreement, Microsoft provided web crawling services by using its web crawler technology to collect web content on OpenAI's behalf for OpenAI to use for training its LLMs. News Ex. 634 at (OPCO_MDL_004771618).

4. Distribution – Taxi and transfer of GPT-3 training data

827. Microsoft and OpenAI copied News Plaintiffs' content to sell and trade that content to and with one another. *See* Section II.C.	827. **Disputed – Not Supported by Cited Evidence; Improper Legal Argument.** Defendants respond that Paragraph 827 is disputed. The cited support references "Section II.C.," which does not itself constitute evidence and spans over one-hundred Paragraphs concerning a breadth of subject matter without reasonable specificity as to what alleged "cop[ying]," "sell[ing]" and/or "trad[ing]" this Paragraph is intended to encompass. For the purposes of this response, Defendants assume that this Paragraph concerns "Taxi and transfer of GPT-3 training data" as described in Paragraphs 694-705 (which are not facts set forth in a discrete subsection), and therefore incorporate by reference their responses to Paragraphs 694-705 in their entirety. To the extent Paragraph 827 is intended to encompass different and/or additional subject matter than that identified herein, Defendants too dispute any such allegations of "cop[ying]," "sell[ing]" and/or "trad[ing]" as Plaintiffs cite no evidence that Microsoft or OpenAI sold or traded Plaintiffs' content to or with one another.

	Defendants further respond that the Court should disregard Paragraph 827 in its entirety because the assertion constitutes improper attorney argument and/or is conclusory.

B. Use of News Content for Data Processing

828. Before training the models at issue, OpenAI processed the acquired data by, *inter alia*, extracting, deduplicating, filtering, and tokenizing it. Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at 14; Ex. 633 [Alex Paino 30(b)(6) Deposition Tr.] at 19:6-20:10); Ex. 521 [Yuan Depo Tr.] at 34:5-7.	828. **Undisputed.** Microsoft defers to OpenAI.
829. *Extraction*: During the "extraction" process, OpenAI extracted the text it wanted to train on from the HTML data that it had acquired. Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at 14.	829. **Disputed in Part – Not Supported by Cited Evidence.** OpenAI responds as follows: Dispute to the extent that it implies OpenAI's text extraction process extracts *only* and *all* the text it wants to train on. Pursuant to Fed. R. Evid. 106, text extraction is a process that relies on heuristics and will often (1) extract text that is undesirable for training and (2) fail to extract text that is desirable for training. *See* News Ex. 1171 (B. Frederiksen-Cross Rebuttal Rep.) ¶¶ 48–51. For that reason, the disputed proposition is not supported by the cited evidence or the record. Microsoft defers to OpenAI.

830. OpenAI performed this extraction using HTML-to-text extractors; the specific HTML-to-text extractors used by OpenAI have varied from model to model. Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at 14.	830. **Undisputed.** Microsoft defers to OpenAI.
831. For GPT-3 OpenAI used publicly available version of HTML-to-text extractors called "newspaper" and "dragnet." Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at 14; Ex. 756 [OpenAI's Supplemental Response to Intercept Interrogatory No. 3 (Dec. 31, 2025)] at 6.	831. **Undisputed.** Microsoft defers to OpenAI.
832. For GPT-3.5, OpenAI developed and used its own implementation of the "newspaper" text extractor. Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at 14; Ex. 756 [OpenAI's Supplemental Response to Intercept Interrogatory No. 3 (Dec. 31, 2025)] at 6.	832. **Undisputed.** Microsoft defers to OpenAI.
833. For GPT-4, OpenAI used an OpenAI-developed HTML-to-text extractor called "gutentag" and the OpenAI-developed implementation of the "newspaper" extractor. Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's	833. **Undisputed.** Microsoft defers to OpenAI.

Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at 14; Ex. 756 [OpenAI's Supplemental Response to Intercept Interrogatory No. 3 (Dec. 31, 2025)] at 6.	
834. For GPT-4o, OpenAI used another version of the OpenAI-developed HTML-to-text extractor called "gutentag." Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at 14; Ex. 756 [OpenAI's Supplemental Response to Intercept Interrogatory No. 3 (Dec. 31, 2025)] at 6.	834. **Undisputed.** Microsoft defers to OpenAI.
835. *Deduplication*: Duplication of data in the training datasets is known to increase the risk of regurgitation. Ex. 1264 [https://openai.com/index/openai-and-journalism/] (OpenAI January 2024 blog post stating that "memorization" is "more common when particular content appears more than once in training data").	835. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 1264 states that "memorization" is "more common when particular content appears more than once in training data." However, disputed to the extent Plaintiffs imply this statement establishes that duplication of training data is "known" to increase the risk of regurgitation which is not supported by the cited evidence. Pursuant to Fed. R. Evid. 106, the cited document discusses memorization, not regurgitation, which are distinct concepts. *See* A. Blanco Opp. Decl. Ex. 250 (A. Paino 30(b)(6) Tr.) 62:2–63:1 (describing many definitions of overfitting and noting that memorization is distinct from overfitting), 63:2–22 (describing relationship between overfitting and regurgitation); A. Blanco Opp.

	Decl. Ex. 250 (A. Paino 30(b)(6) Tr.) 68:15–69:7 (noting that "regurgitation" has many definitions). Microsoft defers to OpenAI.
836. OpenAI did not deduplicate the training datasets for GPT-2 and GPT-3. Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 756 [OpenAI's Supplemental Response to Intercept Interrogatory No. 3 (Dec. 31, 2025)] at 6-7; Ex. 739 [OPCO_MDL_004806459] at –460, -469.	836. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 836 are disputed because they are not supported by the cited evidence. News Exs. 748, 756, and 739 discuss deduplication practices for models developed after GPT-3, but do not establish that deduplication was absent from the GPT-2 or GPT-3 training datasets. Pursuant to Fed. R. Evid. 106, News Ex. 680 is an OpenAI paper stating that OpenAI performed document-level fuzzy deduplication for GPT-3. *See* News Ex. 680 at OPCO_MDL_000000446. Microsoft defers to OpenAI.
837. OpenAI did not comprehensively deduplicate the training datasets for GPT-4, GPT-3.5, GPT-3.5 Turbo, GPT-4.5 Turbo, GPT-4o, and GPT-4o mini. Ex. 521 [Yuan Depo. Tr.] at 62:22-23 (defining fuzzy deduplication as deduplication of documents above some similarity threshold); Ex. 741 [OPCO_MDL_004762295] at -301 (recognizing fuzzy deduplication threshold "is quite high since [it] need[s] a lot of data" and thus "allows many close duplicates in"); Ex. 741 [OPCO_MDL_004762295] at –301 ("the other big issue" with fuzzy deduplication is it is on a per document level); Ex. 739 [OPCO_MDL_004806459] at	837. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 837 are disputed because the cited evidence does not support the claim. The cited evidence does not establish that OpenAI failed to "comprehensively deduplicate" the training datasets for the identified models. Plaintiffs do not define "comprehensive deduplication" or establish that OpenAI failed to meet such a standard. Rather, the cited documents discuss different approaches to deduplication, including document-level

-469 (even document-level dataset deduplication is insufficient for regurgitation prevention).	fuzzy deduplication, and the tradeoffs associated with different similarity thresholds. Microsoft defers to OpenAI.
838. The training datasets for GPT-2, GPT-3, GPT-4, GPT-3.5, GPT-3.5 Turbo, GPT-4.5 Turbo, GPT-4o, and GPT-4o mini contain many duplicate copies of News Plaintiffs' Asserted Works. *See* Facts 1185-86, *infra*.	838. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Plaintiffs cite no specific record evidence to support their contention in Paragraph 838, as required by Fed. R. Civ. P. 56(c)(1). Plaintiffs' general cross-reference to a separate section of their Rule 56.1 statement, without identifying particular evidence supporting this assertion, is insufficient to establish the asserted fact. To the extent Plaintiffs rely on Facts 1185 and 1186, *infra*, Defendants incorporate by reference their responses to those facts. Microsoft defers to OpenAI.
839. *Filtering*: OpenAI filtered Common Crawl data to remove material it deemed low quality and to prioritize higher-quality documents, including by using classifier-based filtering methods. Ex. 748 [OpenAI's Responses to Daily News Plaintiffs' Fifth Set of Interrogatories], No. 17; Ex. 752 [OpenAI's Supplemental Response to Ziff Davis Interrogatory No. 20 (Feb. 27, 2026)] at p. 15.	839. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI used classifier-based filtering methods to process Common Crawl data. However, disputed to the extent Plaintiffs characterize these methods as designed to prioritize "higher-quality documents" because Plaintiffs do not define "higher quality" and the cited evidence does not support that characterization. Specifically, the logistic regression classifier is a classification program that differentiates between (a) unreadable, unusable, repetitive, or machine-generated data and (b) potentially usable language data. Accordingly, and pursuant to Fed. R. Evid.

	106, the classifier's assessment of "quality" concerned whether a document's token distribution was consistent with human-written language, rather than a broader assessment of the substantive quality or value of the content. *See* News Ex. 748, No. 17; News Ex. 752 at 15; News Ex. 483 ¶ 357(e).357(e). Microsoft defers to OpenAI.
840. OpenAI understood that "to improve the quality of Common Crawl" it had to "develop[] an automatic filtering method to remove low-quality documents." Ex. 339 [OPCO_MDL_004794796] at –802.	840. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI used classifier-based filtering methods to process Common Crawl data. However, disputed to the extent Plaintiffs characterize these methods as designed to prioritize "higher-quality documents" because Plaintiffs do not define "higher quality" and the cited evidence does not support that characterization. Specifically, the logistic regression classifier is a classification program that differentiates between (a) unreadable, unusable, repetitive, or machine-generated data and (b) potentially usable language data. Accordingly, and pursuant to Fed. R. Evid. 106, the classifier's assessment of "quality" concerned whether a document's token distribution was consistent with human-written language, rather than a broader assessment of the substantive quality or value of the content. *See* News Ex. 748, No. 17; News Ex. 752 at 15; News Ex. 483 ¶ 357(e). Microsoft defers to OpenAI.

841. To "improve the quality of Common Crawl," OpenAI trained a logistic regression classifier using curated datasets like WebText "as a proxy for high-quality documents." Ex. 339 [OPCO_MDL_004794796] at –802.	841. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI used classifier-based filtering methods to process Common Crawl data. However, disputed to the extent Plaintiffs characterize these methods as designed to prioritize "higher-quality documents" because Plaintiffs do not define "higher quality" and the cited evidence does not support that characterization. Specifically, the logistic regression classifier is a classification program that differentiates between (a) unreadable, unusable, repetitive, or machine-generated data and (b) potentially usable language data. Accordingly, and pursuant to Fed. R. Evid. 106, the classifier's assessment of "quality" concerned whether a document's token distribution was consistent with human-written language, rather than a broader assessment of the substantive quality or value of the content. *See* News Ex. 748, No. 17; News Ex. 752 at 15; News Ex. 483 ¶ 357(e). Microsoft defers to OpenAI.
842. The New York Times Annotated Corpus was another curated dataset used as a proxy for high-quality documents for training the logistic regression classifier. *See infra* Facts 1093-1136 (discussing The New York Times Annotated Corpus).	842. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI used classifier-based filtering methods to process Common Crawl data. However, disputed to the extent Plaintiffs characterize these methods as designed to prioritize "higher-quality documents" because Plaintiffs do not define "higher quality" and the cited evidence does not support that characterization. Specifically, the logistic

	regression classifier is a classification program that differentiates between (a) unreadable, unusable, repetitive, or machine-generated data and (b) potentially usable language data. Accordingly, and pursuant to Fed. R. Evid. 106, the classifier's assessment of "quality" concerned whether a document's token distribution was consistent with human-written language, rather than a broader assessment of the substantive quality or value of the content. *See* News Ex. 748, No. 17; News Ex. 752 at 15; News Ex. 483 ¶ 357(e). *See also infra* Counterstatement to Facts 1093-1136 (discussing The New York Times Annotated Corpus). Microsoft defers to OpenAI.
843. OpenAI used news content as a part of its filtering process for identifying quality data for training; for example, OpenAI training a logistic regression filtering using Webtext, Wikipedia, the New York Times Annotated Corpus, and Libgen as the "positive class." Ex. 743 [OPCO_MDL_000479924] at –925; Ex. 621 [Ryder 30(b)(6) Deposition Tr.] at 118:9-24.	843. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that OpenAI used classifier-based filtering methods to process Common Crawl data. However, disputed to the extent Plaintiffs characterize these methods as designed to prioritize "quality data" without defining "quality data." Specifically, the logistic regression classifier is a classification program that differentiates between (a) unreadable, unusable, repetitive, or machine-generated data and (b) potentially usable language data Accordingly, the classifier's assessment of "quality" concerned whether a document's token distribution was consistent with human-written language, rather than a broader assessment of the substantive quality or value of the content. *See* News

	Ex. 748, No. 17; News Ex. 752 at 15; News Ex. 483 ¶ 357(e). The disputed proposition is therefore not supported by the cited evidence. Microsoft defers to OpenAI.
844. The "[t]op 100 domains by document count for [redacted] (i.e., the filtered Common Crawl dataset OpenAI used to train [redacted] includes "chicagotribune.com" (2,182,079 documents) and "nytimes.com" (2,064,805 documents), among other news publishers. Ex. 493 [OPCO_MDL_00133274] at –274 and –275.	844. **Disputed in Part – Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that the quoted text appears in the cited document. News Ex. 493 at OPCO_MDL_00133274–275. However, disputed to the extent that this implies the top 100 domains are all news publishers; they are not. *See* News Ex. 493 at OPCO_MDL_00133274 (showing indeed.com, google.com, imdb.com, stackexchange.com, groupon.com, and many other non-news domains). Microsoft defers to OpenAI.
845. *Tokenization*: Before training, OpenAI tokenized the training data. Ex. 633 [Paino 30(b)(6)] at 20:7-15; Ex. 529 [Chess Deposition Tr.] at 31:1-11.	845. **Undisputed.** Microsoft defers to OpenAI.
846. Tokenization refers to the process of converting raw text into discrete "tokens" that each correspond to a numerical identifier. Ex. 633 [Paino 30(b)(6)] at 20:7-15; Ex. 529 [Chess Deposition Tr.] at 31:1-11.	846. **Undisputed.**

C. Use of News Content During Training

847. The at-issue LLMs contain billions of parameters (weights) used to generate outputs (responses) to inputs (prompts) one token (i.e., a unit of text) at a time. Ex. 461 [Berg-Kirkpatrick Opening ¶ 30]; Ex. 483 [Callison-Burch Opening Report ¶ 138-40].	847. **Undisputed.**

1. Use of News Content During Pre-Training

848. Pre-training is the initial stage of model training, in which the model is trained on a large training corpus. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 33.	848. **Undisputed.**
849. During pre-training, OpenAI created and passed millions of copies of News Plaintiffs' articles through the model. Ex. 676 [Callison-Burch Rebuttal Rep.] ¶ 59 ("Again, each of these steps requires creating and storing copies of parts of the data in ephemeral memory."); *see also* Facts 1184-1187, 1210-12, 1221-1223, *infra*.	849. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The cited evidence does not establish that, during pre-training, OpenAI "created and passed millions of copies of News Plaintiffs' articles through the model." . . . Dr. Callison-Burch states generally that processing training data can create copies but does not quantify those copies or state that millions of copies of News Plaintiffs' articles were created during training. To the extent this statement relies on Plaintiffs' separate assertions concerning millions of training-data records that allegedly match News Plaintiffs' works, OpenAI incorporates, as applicable, its responses to Facts 1184–1187, 1210–1212, and 1221–1223. Microsoft defers to OpenAI.

850. OpenAI's expert, Dr. Callison-Burch, explained that when developing an LLM, "having data in a particular domain you care about is often useful [for training]." Ex. 635 [Callison-Burch Depo. Tr.] at 230:17-22.	850. **Undisputed.** Microsoft defers to OpenAI.
851. Dr. Callison-Burch agreed that if the goal was to create an LLM that could answer queries about certain events, a developer would want to train the model on documents covering those events. Ex. 635 [Callison-Burch Depo. Tr.] at 235:9-19.	851. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed to the extent that Plaintiffs' characterization omits the full context of Dr. Callison-Burch's testimony, which was cabined to certain models: "Yes, assuming that you have a model that's fixed. So this would be not necessarily the models in questions with their full set of capabilities. But a fixed model with only access to pretraining data, then yes." *See* A. Blanco Opp. Decl. Ex. 181 (C. Callison-Burch Tr. (July 15, 2026)) 235:9-236:9. Microsoft defers to OpenAI.
852. Dr. Callison-Burch conceded that a model trained on news content would generate outputs influenced by that content. Ex. 635 [Callison-Burch Depo. Tr.] at 233:2-234:8.	852. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that, in response to Plaintiffs' question "And the data that the model sees during training would influence the outputs?" Dr. Callison-Burch answered "Correct." Disputed that Dr. Callison-Burch "conceded" this point. *See* ECF 1748 (C. Callison-Burch Decl.) § VI ("OpenAI's LLMs Require a Massive Amount of Training Data"); CCB Opp. Decl. (citing CCB Rebuttal Report Section V (plaintiffs' experts overstate the value of their clients' content to LLM training)).

	Microsoft defers to OpenAI.
853. OpenAI decided what training data went into the at-issue models. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 17:15-25.	853. **Undisputed.** Microsoft defers to OpenAI.
854. During training, the LLM is provided with a sequence of tokens (e.g., words) from its training data and the LLM calculates a probability distribution over the potential next tokens in that sequence. Goldstein Opening Rep. n. 146; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 160 ("'Train loss' generally measures the model's ability to predict words in text sequences drawn from the 'training' dataset"), 181.	854. **Undisputed.**
855. With each training pass, the LLM's calculated probability distribution is compared to the "correct" distribution in which the actual next token in the training data sequence has high probability. *See* [Goldstein Opening Rep.]; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 160, 181 ("The model's prediction was then compared against the true next token from the dataset").	855. **Undisputed.**
856. The procedure for comparing the output probability distribution with the correct one is called "computing the loss function." *See* Goldstein Opening Rep. n. 146; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 160, 181.	856. **Disputed in Part**. OpenAI responds as follows: Undisputed as to Plaintiffs' statement that there is a procedure for comparing the output probability distribution with a reference value. ECF 1819-1 (T. Goldstein Decl. Ex. A) ¶ 125, n. 146; News Ex. 483 (C. Callison-Burch Opening Rep.) ¶¶ 160, 181. However, disputed as to Plaintiffs' characterization as vague and imprecise because they do not distinguish between "train

	loss" and "test loss," and that characterization is not supported by the cited evidence. Specifically, train loss refers to the loss function in the training context, i.e., measuring the model's ability to predict words in text sequences drawn from the model's training dataset (the data set aside for training). There is also a separate loss function, test loss (or validation loss), which measures the model's ability to accurately predict missing words in text sequences drawn from the test dataset, i.e., the data set aside for testing. *See* ECF 1748 (C. Callison-Burch Decl.) ¶¶ 31, 46. Microsoft defers to OpenAI.
857. The "loss function" is a mathematical formula, defined by the LLM's human designers, that measures how close the LLM's output and "correct" probability distributions are (i.e., the "loss"). *See* Goldstein Opening Rep. n. 146; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 160, 181 ("This comparison yielded a numerical measure of error, called the 'loss,' which quantified the gap between what the model predicted and what actually appeared in the text").	857. **Disputed in Part**. OpenAI responds as follows: Undisputed that "loss function" is a mathematical formula that measures how close the probability distributions generated by the LLM are to a reference value. *But see* response to Paragraph 856 (discussing the distinction between train loss and test loss). *See* ECF 1748 (C. Callison-Burch Decl.) ¶¶ 31, 46. Microsoft defers to OpenAI.
858. To pretrain the at-issue models, the models were optimized against a selected "loss function." Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 175:18-19].	858. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: The facts set forth in Paragraph 858 are disputed in part because they are not supported by the cited evidence. Specifically, Plaintiffs' "models were optimized against" characterization is unsupported by the cited evidence. Mr. Ryder testified that

	"[f]or all the models we are talking about, they were trained with next token prediction loss." Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr.) 175:22-176:1. The entirety of the cited quote is "And the thing you want to optimize against is often called the loss function." *Id.* 175:18-19. Microsoft defers to OpenAI.
859. OpenAI pretrained the models at issue in this case using next-token prediction loss. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 175:25-176:1; Ex. 800 [Lafferty Rebuttal Rep.] ¶ 21.	859. **Undisputed.** Microsoft defers to OpenAI.
860. Next-token prediction loss measures how well the model predicts the next token in text from its training samples. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 177:11-14, 178:9-14; Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 33.	860. **Undisputed.**
861. By optimizing next-token prediction loss during training, OpenAI calibrated its models model to accurately predict the next word in a sequence from the training data. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 175:10-21, 178:9-14. Ex. 483 [Callison-Burch Opening Rep.] ¶ 396 ("OpenAI's models are 'pre-trained to predict the next token in a document.'").	861. **Disputed in Part**. OpenAI responds as follows: Undisputed that OpenAI tries to optimize next-token prediction loss during training, but the objective is for models to generalize. Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr.) 175:10–21, 178:9–14; News Ex. 483 (Callison-Burch Opening Rep.) ¶ 396. However, disputed as to the characterization that OpenAI calibrated its models to accurately predict the next word in a sequence from the training data because that characterization is not supported by the cited evidence. Pursuant to Fed. R. Evid. 106, the models are designed to assign a higher probability to the actual next token in the training data; they predict

	over all of the training data, and the primary measure of the prediction is loss on the test data. A. Blanco Opp. Decl. Ex. 250 (A. Paino 30(b)(6) Tr.) 61:9–13; Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr.) 21:1–23; News Ex. 517 (J. Lafferty Opening Rep.) ¶ 21. Microsoft defers to OpenAI.
862. Next-token prediction loss is a measure of model performance. Ex. 483 [Callison-Burch Opening Rep.] ¶ 310; *See also* Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 178:15-18.	862. **Disputed in Part - Not Supported by Cited Evidence.** Defendants respond as follows: The facts set forth in Paragraph 862 are disputed in part because they are not supported by the cited evidence. Specifically, disputed that the cited evidence supports the contention that next-token prediction loss is a measure of overall model performance. Mr. Ryder testified that one way to think of next-token prediction objective is looking at cross entropy and next-token prediction loss, but he does not suggest that this is about model performance. Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr.) 178:15–18. Dr. Callison-Burch's statement that "cross-entropy loss" can be a "rough measure of model performance" was made in the context of evaluating performance on the next-token prediction objective, not overall model quality or performance on downstream tasks. Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 310. Pursuant to Fed. R. Evid. 106, *see* A. Blanco Opp. Decl. Ex. 250 (A. Paino 30(b)(6) Tr.) 61:9–13; Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr.) 21:1–23.
863. OpenAI's technical expert Dr. Chris Callison-Burch opines that, for the "LLMs at issue" in this case,"	863. **Disputed in Part - Incomplete.** OpenAI responds as follows: Undisputed that Dr. Callison-Burch's

"[n]o individual work or set of works is required" for training and that "any individual source" is "entirely fungible." Ex. 483 [Callison-Burch Opening Rep.] ¶ 328.	report contains the cited text. However, pursuant to Fed. R. Evid. 106, Dr. Callison-Burch's full statement says: "[I]n a large-scale pre-training operation, what matters is the composition of the training corpus as a whole, primarily because the purpose of pre-training is to identify broad patterns that manifest in numerous different examples and contexts. No individual work or set of works is required to achieve a diverse distribution of data; indeed, any individual source, particularly in a given domain, is entirely fungible." *See* News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 328). Further, Dr. Callison-Burch opined that LLMs' success "in no way depends on the inclusion of any specific document or group of documents in the training set." News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 344. And, Dr. Callison-Burch opined that "[p]articularly critical here is diversity: to create an LLM capable of comprehending and responding to the full scope of possible user queries and use-cases, OpenAI needed a dataset that included near-universal coverage across every possible dimension, including subject matter, sources, forms of language (formal, colloquial), styles of writing, etc." *Id.* ¶ 329. Dr. Callison-Burch also opined that "[t]o the extent any of the plaintiffs' works were part of that massive and diverse corpus of natural language text used to pretrain the disputed OpenAI LLMs and products, it was incidental to, and solely to serve, this new and different purpose." *Id.* ¶ 327.

	Microsoft defers to OpenAI.
864. OpenAI's expert Dr. Avi Goldfarb opines that "individual news sources are largely interchangeable within the enormous corpus of data used for LLM training." Ex. 498 [Goldfarb Reply Rep.] ¶ 267.	864. **Undisputed.** Microsoft defers to OpenAI.
865. OpenAI's damages expert Dr. Stamm opines that "News Plaintiffs' Asserted Works were substitutable with that of other publishers for purposes of LLM training." Ex. 128 [Stamm Rebuttal Rep.] ¶ 308; Ex. 129 [Stamm ZD Rebuttal Rep.] at ¶ 111.	865. **Undisputed.** Microsoft defers to OpenAI.
866. Microsoft's technical expert Dr. John Lafferty opines: "[N]o individual text or category of texts within a training corpus of this scale is necessary to the model's capability" and "the removal of any single work or collection of works from a corpus of trillions of tokens would not materially alter the statistical structure the model learns." Ex. 517 [Lafferty Opening Rep.] ¶ 87.	866. **Undisputed**. OpenAI defers to Microsoft's response.
867. Dr. Lafferty testified that news content "is not necessary" to the models' performance because "engineers, model developers, would have ways of training the model in a different fashion to compensate for the lack of a very, very broad source such as news data." Ex. 518 [Lafferty Depo. Tr.] at 51:17-23.	867. **Undisputed - Incomplete.** Microsoft responds as follows: Microsoft does not dispute that Dr. Lafferty made the statements in Paragraph 867. But pursuant to Fed. R. Evid. 106, Dr. Lafferty opines that an LLM's exposure to news-like text may improve a model's performance; news does not have a unique functional importance to general LLM capability because the distributional properties of LLM outputs are derived from properties of text used in pre-training (and to be more

	useful, a model should be trained on a large, diverse set of texts). *See* News Ex. 800 (J. Lafferty Rebuttal Rep.) ¶¶ 59, 69–72, 74-78; ECF 1723 (J. Lafferty Decl.) ¶¶ 34, 41, 94, 97. OpenAI defers to Microsoft's response.
868. OpenAI considered news content high quality data for training its LLMs. Ex. 519 [Brundage Depo. Tr.] at 166:22-167:6 ("in general, when you're training language models, you want to have high quality data" and that "newspapers seem obviously like among the kinds of high quality data sources one would consider [] when training language models.").	868. **Disputed in Part.** OpenAI responds as follows: Undisputed that the fact is correct as stated. A. Blanco Opp. Decl. Ex. 177 (M. Brundage 30(b)(1) Tr.) 166:22–167:6. However, disputed that the fact accurately reflects the specific type of quality that news content provides for training, because Plaintiffs' characterization overstates the cited evidence. Specifically, Mr. Brundage explains that [redacted]" A. Blanco Opp. Decl. Ex. 177 (M. Brundage 30(b)(1) Tr.) 166:20-167:6. Further disputed on the basis that former employee Mr. Brundage does not speak on behalf of OpenAI. Microsoft defers to OpenAI.
869. An OpenAI internal document [redacted]. Ex. 520 [OPCO_MDL_001894242]; Ex. 521 [Yuan Depo. Tr.] at 124:14-19.	869. **Disputed in Part.** OpenAI responds as follows: Undisputed that the document contains the quoted language. However, disputed that the fact, as written, is supported by the cited evidence, because it is misleading as [redacted]. Further disputed because the cited deposition testimony is not

	referring to the OpenAI internal document being referenced. A. Blanco Opp. Decl. Ex. 291 (Q. Yuan 30(b)(1) Tr.) 124:14–19. Microsoft defers to OpenAI.
870. OpenAI specifically targeted The Times's sites for training purposes. Ex. 130 [Brockman Depo. Tr.] at 345:25-346:2. (Mr. Brockman agreeing that "[OpenAI] designated high-quality sites" and "New York Times was among them.").	870. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 870 are disputed because they are not supported by the cited evidence. The cited testimony does not support the statement that OpenAI specifically targeted The Times's sites for training purposes; rather, the portion of testimony cited is a question from Plaintiffs' counsel asking, "However you designated high-quality sites, New York Times was among them, in your mind, correct?" *See* A. Blanco Opp. Decl. Ex. 176 (G. Brockman 30(b)(1) Tr.) 345:25–346:2). Microsoft defers to OpenAI.
871. OpenAI sampled data sets that it views as high quality most frequently, and WebText2 was among the datasets sampled 2-3 times during training. Ex. 680 [OPCO_MDL_000000439] at -446-47.	871. **Undisputed.** Microsoft defers to OpenAI.
872. OpenAI's models are "excellent at news" and "extremely good at predicting what comes next in random NYT articles." Ex. 745 [OPCO_MDL_000220773] at 783-784 ("excellent at news"); Ex. 747 [OPCO_MDL_001900258] at -270.	872. **Disputed in Part.** OpenAI responds as follows: Undisputed that the documents state the language quoted. However, disputed to the extent Plaintiffs suggest that the LLMs in dispute generate "copies" of the Asserted Works at issue in this Motion. Pursuant to Fed. R. Evid.

	106, Greg Brockman, who wrote the relevant messages, explained that the Davinci model was "very good at any news task" because a user could provide the model with "context - a document [it has] never seen before, totally new - and [it does not] output a copy." Ex. 176 (G. Brockman 30(b)(1) Tr.) 289:6–24. Mr. Brockman further noted that the model was "[s]imilar to how a phone can have auto-complete" because "[i]t's never going to get it perfectly right, never going to be a copy." *Id.* 289:25–290:18. Mr. Brockman also noted that the model "had not been trained" on the New York Times articles referenced in his message. *Id.* 268:2–8. Microsoft defers to OpenAI.
873. Including news content in OpenAI's LLM pretraining data corpus improved its models' ability to produce news-style outputs. Wenger Opening Rep. at 16; Ex. 800 [Lafferty Rebuttal Rep.] ¶ 69 (noting that "exposure to news-like or book-like text may improve a model's ability to generate news-like or book-like outputs"); Ex. 676 [Callison-Burch Rebuttal Rep.] ¶ 35 (noting the at-issue LLMs "may generate text that is 'similar to' text in the training data"); Ex. 130 [Brockman Tr.] at 300:15-18 ("[W]hen you train these models on data for – they're relevant to specific areas, broad areas, like coding, they get better at that area."); Ex. 749 [OPCO_MDL_004738943] at -43 (OpenAI internal document stating that "[w]hen thinking about whether or not to include a new dataset in	873. **Disputed in Part.** OpenAI responds as follows: Undisputed that the documents say what is cited. However, disputed to the extent that the quotes are incomplete and missing context. The pincite to the E. Wenger Opening Report is incorrect. Further, the Lafferty Rebuttal Report parenthetical is incomplete. Pursuant to Fed. R. Evid. 106, when discussing Dr. Etzioni's and Dr. Wenger's Opening Reports, Dr. Lafferty says: "At most, these statements suggest that exposure to news-like or book-like text may improve a model's ability to generate news-like or book-like outputs." News Ex. 800 (J. Lafferty Rebuttal Rep.) ¶ 69. The parenthetical for Dr. Callison-Burch's Rebuttal Report is

our pre-training runs, there should be only two considerations," including whether "the capability that it advances [is] something we want"); Ex. 1137 [https://nonint.com/2023/06/10/the-it-in-ai-models-is-the-dataset/] (October 2023 blog post published by OpenAI employee James Betker: "It's becoming awfully clear to me that these models are truly approximating their datasets to an incredible degree"); *see also supra* Ex. 745 [OPCO_MDL_000220773] at 783-784 ("excellent at news"); Ex. 747 [OPCO_MDL_001900258] at -270.	accurate, but he too includes a note specifying that the models do not specifically learn to mimic the specific news they were trained on. News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶ 35. The parenthetical for Mr. Brockman's testimony is also incomplete. Pursuant to Fed. R. Evid. 106, immediately after the quoted testimony, Mr. Brockman clarifies: "I think we have less results for very specific domains. And I kind of consider news to be more of a specific as opposed to English language, German language, those kinds of broad ones." Ex. 176 (G. Brockman 30(b)(1) Tr.) 300:19–23. After the deposing attorney tried to push this opinion on him again, Mr. Brockman again pushed back. *Id.* 300:24–302:17. Microsoft defers to OpenAI.
874. According to OpenAI's LLM-training expert, the success of OpenAI's models does not depend on the inclusion of any one document or group of documents in the training set. Ex. 483 [Callison-Burch Opening Rep.] ¶ 344.	874. **Undisputed**. Microsoft defers to OpenAI.
875. OpenAI's expert, Dr. Callison-Burch, opined that OpenAI could have attained its training objectives without using the copies it made of News Plaintiffs' Asserted works. Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 328, 343, 344; Ex. 676 [Callison-Burch Rebuttal Rep.] ¶ 50.	875. **Disputed - Not Supported by Cited Evidence**. OpenAI responds as follows: The facts set forth in Paragraph 875 are disputed because they are not supported by the cited evidence. Specifically, the cited evidence offered by Plaintiffs broadly opines that the inclusion or exclusion of any one document (not the "News Plaintiffs' Asserted Works" at issue in this Motion) within a model's training data does not determine the model's success. News Ex. 483 (C. Callison-Burch Opening Rep.) ¶¶

	328, 343, 344; News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶ 50. While Dr. Callison-Burch said that no one's work is necessary for training, and the data is fungible, he does not make any claims specifically as to Plaintiffs' Asserted Works in these cited paragraphs. *Id*. Dr. Callison-Burch's opinion is that a massive corpus of training data is needed for training models. Microsoft defers to OpenAI.

2. Use of News Content During Mid-Training

876. Mid-training refers to continued training of a pre-trained model. Ex. 621 [Ryder 30(b)(6) Deposition Tr.] at 139:1-4.	876. **Undisputed.**
877. The technical methods used for mid-training are similar to those used during pre-training. Ex. 637 [Zoph 30(b)(6) Depo. Tr.] at 237:8-11; Ex. 483 [Callison-Burch Opening Rep.] ¶ 259.	877. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that the cited evidence supports the broad assertion that "[t]he technical methods used for mid-training are similar to those used during pre-training," as stated. Mr. Zoph's cited testimony states only that mid-training is closer to pre-training than post-training, and the cited excerpts of Dr. Callison-Burch's report do not compare "technical methods" used by mid-training and pre-training. Pursuant to Fed. R. Evid. 106, OpenAI's corporate designee Chris Koch testified that mid-training "can encompass a lot of other things as well [as the normal pre-training data intervention purposes], like architecture work, or various

	other types of improvements that would want to make to the model that are different than data interventions." *See* Ex. 223 (C. Koch 30(b)(6) Tr.) 21:3–6. Additionally disputed because the statement itself is overly reductive, as mid-training encompasses a number of different techniques (some of which may resemble pre-training). Microsoft defers to OpenAI.
878. OpenAI mid-trained [redacted] *See* Ex. 751 [OPCO_MDL_004738868] at 887-893. *See* Ex. 1138 [OpenAI's Supp. Response to News Plaintiffs' Joint RFA No. 275].	878. **Undisputed.** Microsoft defers to OpenAI.
879. GPT-4 Turbo was "continued training of GPT-4." Ex. 633 [Paino 30(b)(6) Depo. Tr.] at 114:21-25.	879. **Undisputed.** Microsoft defers to OpenAI.
880. An internal OpenAI presentation describes GPT-4 Turbo as a "finetune of GPT-4" which, *inter alia*, "updates the knowledge cutoff (from September 2021 to April 2023)." Ex. 753 [OPCO_MDL_000155665] at 3.	880. **Undisputed.** Microsoft defers to OpenAI.
881. During mid-training, developers may adjust the data mixture, emphasize different sources, introduce domain-specific corpora, increase context length, improve freshness, or improve instruction-following. Ex. 483 [Callison-Burch Opening Rep.] ¶ 259.	881. **Undisputed.**

882. One goal of mid-training is to make the model "significantly more intelligent on an axis [OpenAI] cares about." Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 138:7-9.	882. **Undisputed.** Microsoft defers to OpenAI.
883. OpenAI also used mid-training to update the model's knowledge with news and events taking place after the pretraining dataset was assembled (i.e., "fresh" data). Ex. 755 [OPCO_MDL_004585389] at 8; Ex. 753 [OPCO_MDL_000155665] at 6.	883. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 883 are disputed because they are not supported by the cited evidence. Disputed that the cited evidence supports Plaintiffs' characterization that "fresh data" or "freshness" is limited to only news or current events. Pursuant to Fed. R. Evid.106, mid-training included many different types of data generated after the pretraining datasets were assembled (i.e., "up-to-date world knowledge") beyond news. Ex. 223 (C. Koch 30(b)(6) Tr.) 21:13–21. Further disputed to the extent Plaintiffs imply that news-related data is particularly valuable for mid-training; to the contrary, and pursuant to Fed. R. Evid. 106, Dr. Callison-Burch discusses how "OpenAI's witnesses testified that efforts to use news-specific datasets for mid-training to improve 'freshness' failed—illustrating that news articles are not, in fact, valuable for mid-training." *See* News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶¶ 127–129; *see also* Ex. 223 (C. Koch 30(b)(6) Tr.) 34:20–23. Finally, pursuant to Fed. R. Evid. 106, OpenAI mid-trained on only two models in dispute and then stopped. News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶ 128. Microsoft defers to OpenAI.

884. OpenAI used curated datasets, including datasets targeting news content, during mid-training to enhance the performance of [redacted] *See* Ex. 751 [OPCO_MDL_004738868] at 887-893. *See* Ex. 1138 [OpenAI's Supp. Response to News Plaintiffs' Joint RFA No. 275]; Ex. 639 [Koch 30(b)(6) Depo. Tr.] at 15:2-20, 22:5-8; Ex. 757 [OPCO_MDL_000155738] at 5-6, 10-14; Ex. 755 [OPCO_MDL_004585389] at 8, 17-19.	884. **Disputed in Part.** OpenAI responds as follows: Plaintiffs' misleading characterization of OpenAI's mid-training datasets, and inferences drawn therefrom, are contradicted by record evidence. Pursuant to Fed. R. Evid. 106, OpenAI's corporate designee Chris Koch testified that news content constituted a small percentage of the mid-training mix and that the use of news-specific datasets during mid-training did not enhance the performance of [redacted] Ex. 223 (C. Koch 30(b)(6) Tr.) 21:13–21, 22:5–8, 22:12–17, 23:17–24:336:24–37:1, 37:11–12, 39:10–13, 39:17–21, 40:10–17, 47:2–4, 47:25–48:1. Further disputed to the extent Plaintiffs suggest OpenAI used curated news-specific datasets to mid-train the [redacted]. Pursuant to Fed. R. Evid.106, Dr. Callison-Burch discusses how "OpenAI's witnesses testified that efforts to use news-specific datasets for mid-training to improve 'freshness' failed—illustrating that news articles are not, in fact, valuable for mid-training." News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶¶ 127–129; *see also* Ex. 223 (C. Koch 30(b)(6) Tr.) 34:20–23. Microsoft defers to OpenAI.
885. During "mid-training" OpenAI wanted to improve on "current events / news articles (NYT data, etc.)." Ex. 759 [OPCO_MDL_004740422] at -422.	885. **Disputed in Part.** OpenAI responds as follows: Disputed that the cited source supports the broad contention about OpenAI's goals for mid-training as implemented. Further disputed as to Plaintiffs' characterization of the document to the extent it suggests

	"news articles" were uniquely valuable to mid-training, as that inference is contradicted by record evidence. Pursuant to Fed. R. Evid. 106, OpenAI's corporate designee Chris Koch testified that news content constituted a small percentage of the mid-training mix and that the use of news-specific datasets during mid-training did not enhance the performance of [redacted]. Ex. 223 (C. Koch 30(b)(6) Tr.) 21:13–21, 22:5–8, 22:12–17, 23:17–24:3, 36:24–37:1, 37:11–12, 39:10–13, 39:17–21, 40:10–17, 47:2–4, 47:25–48:1. Further disputed to the extent Plaintiffs suggest OpenAI used curated news-specific datasets to mid-training the [redacted]. Dr. Callison-Burch discusses how "OpenAI's witnesses testified that efforts to use news-specific datasets for mid-training to improve 'freshness' failed—illustrating that news articles are not, in fact, valuable for mid-training." News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶¶ 127–129; *see also* Ex. 223 (C. Koch 30(b)(6) Tr.) 34:20–23. Microsoft defers to OpenAI.
886. Members of OpenAI's mid-training team had the goal to "crush freshness in the domain of real-world news first" with [redacted] accuracy on current events." Ex. 761 [OPCO_MDL_000481870] at -870; Ex. 639 [Koch 30(b)(6) Depo. Tr.] at 33:21-35:18, 38:6-18.	886. **Disputed.** OpenAI responds as follows: This fact is disputed. As confirmed in the cited testimony of OpenAI's corporate representative, Chris Koch, the quoted language reflects only the views of an individual researcher and cannot support a contention that all "[m]embers of OpenAI's mid-training team" shared the same goal as stated. Further disputed to the extent Plaintiffs seek to

	draw inferences from these statements that OpenAI attributed specific value to news articles during mid-training. Additionally, pursuant to Fed. R. Evid. 106, OpenAI's corporate designee Chris Koch testified that news content constituted a small percentage of the mid-training mix and that the use of news-specific datasets during mid-training did not enhance the performance of [redacted]. Ex. 223 (C. Koch 30(b)(6) Tr.) 21:13–21, 22:5–8, 22:12–17, 23:17–24:3, 36:24–37:1, 37:11–12, 39:10–13, 39:17–21, 40:10–17, 47:2–4, 47:25–48:1. Further disputed to the extent Plaintiffs suggest OpenAI used curated news-specific datasets to mid-training the [redacted]. Dr. Callison-Burch discusses how "OpenAI's witnesses testified that efforts to use news-specific datasets for mid-training to improve 'freshness' failed—illustrating that news articles are not, in fact, valuable for mid-training." News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶¶ 127–129; *see also* Ex. 223 (C. Koch 30(b)(6) Tr.) 34:20–23. Microsoft defers to OpenAI.
887. OpenAI had believed that using news data in mid-training would improve freshness on what users care about. Ex. 639 [Koch 30(b)(6) Depo. Tr.] at 34:16-23.	887. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 887 are disputed because it is not supported by the cited testimony. Pursuant to Fed. R. Evid. 106, Chris Koch did not testify as to OpenAI's belief at any point in time. Rather, he acknowledged the existence of "a hypothesis that using news data would be able to improve

	freshness" and he further testified "[t]hat hypothesis later ended up being incorrect." *See* Ex. 223 (C. Koch 30(b)(6) Tr.) 33:21–35:18; *see also id.* 38:6–18. Microsoft defers to OpenAI.
888. OpenAI identified news datasets, including CC-News, as "natural sources of freshness" and "high quality (and fresh) data relevant for our users." Ex. 757 [OPCO_MDL_000155738 at 6]; Ex. 755 [OPCO_MDL_004585389] at 8.	888. **Disputed in Part.** OpenAI responds as follows: Undisputed that CC-News is identified in the cited presentations as one, among many, sources of fresh data. However, disputed to the extent Plaintiffs characterize the documents as evidence that news datasets contain uniquely "highly quality (and fresh) data relevant to the users," as that characterization is contradicted by record evidence. Pursuant to Fed. R. Evid. 106, OpenAI's corporate designee Chris Koch testified that news content constituted a small percentage of the mid-training mix and that the use of news-specific datasets during mid-training did not enhance the performance of [redacted]. Ex. 223 (C. Koch 30(b)(6) Tr.) 21:13–21, 22:5–8, 22:12–17, 23:17–24:3, 36:24–37:1, 37:11–12, 39:10–13, 39:17–21, 40:10–17, 47:2–4, 47:25–48:1. Further disputed to the extent Plaintiffs suggest OpenAI used curated news-specific datasets to mid-training the [redacted]. Dr. Callison-Burch discusses how "OpenAI's witnesses testified that efforts to use news-specific datasets for mid-training to improve 'freshness' failed—illustrating that news articles are not, in fact, valuable for mid-training." News Ex. 676 (C. Callison-Burch Rebuttal Rep.)

	¶¶ 127–129; *see also* Ex. 223 (C. Koch 30(b)(6) Tr.) 34:20–23. Microsoft defers to OpenAI.
889. CC-News is a continuously crawled sub-set of Common Crawl's main crawl released daily with a focus on news websites and other sources of fresh data. Goldstein Opening Rep. ¶ 172.	889. **Undisputed.** Microsoft defers to OpenAI.
890. OpenAI recognized that, for mid-training data, it "want[ed] some level of memorization so that it's useful for freshness." Ex. 763 [OPCO_MDL_001669609] at -609.	890. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that Exhibit 763 contains the quoted language. Disputed that the quoted language, which Plaintiffs misleadingly quote out of context from a technical discussion of specific settings for OpenAI's anti-regurgitation masking, supports the statement as stated. Pursuant to Fed. R. Evid. 106, the full sentence states: "It's kind of a tension between we want some level of memorization so that it's useful for freshness but we don't too much that it becomes an antiregurgitation issue." News Ex. 763. Further disputed to the extent Plaintiffs characterize "memorization" in this document as suggesting that OpenAI's models memorize entire news articles (as opposed to facts or other non-protectable information contained therein). Microsoft defers to OpenAI.
891. OpenAI performed multiple mid-training training runs using news content and deployed models that	891. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth

were mid-trained using news content. Ex. 639 [Koch 30(b)(6) Depo. Tr.] at 29:14-30:3, 36:13-14, 39:15-16.	in Paragraph 891 are disputed because the cited evidence does not support Plaintiffs' characterization that OpenAI performed multiple mid-training runs, and deployed mid-trained models, using only news content. Pursuant to Fed. R. Evid.106, OpenAI's corporate representative Chris Koch testified that "OpenAI deployed multiple mid-trained models, probably two or three . . . that were trained on [redacted] CC-NEWS [redacted]." Ex. 223 (C. Koch 30(b)(6) Tr.) 40:10–13. Further, and pursuant to Fed. R. Evid. 106, Mr. Koch also testified that news content constituted a small percentage of the mid-training mix and that the use of news-specific datasets during mid-training did not enhance the performance of [redacted]. Ex. 223 (C. Koch 30(b)(6) Tr.) 21:13–21, 22:5–8, 22:12–17, 23:17–24:3, 36:24–37:1, 37:11–12, 39:10–13, 39:17–21, 40:10–17, 47:2–4, 47:25–48:1. Further disputed to the extent Plaintiffs suggest OpenAI used curated news-specific datasets to mid-training the [redacted]. Dr. Callison-Burch discusses how "OpenAI's witnesses testified that efforts to use news-specific datasets for mid-training to improve 'freshness' failed—illustrating that news articles are not, in fact, valuable for mid-training." News Ex. 676 (C. Callison-Burch Rebuttal Rep.) ¶¶ 127–129; *see also* Ex. 223 (C. Koch 30(b)(6) Tr.) 34:20–23. Microsoft defers to OpenAI.

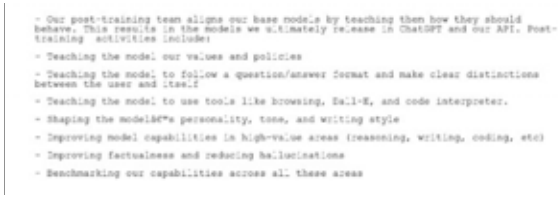
892. The CC-News mid-training dataset [redacted] contained records matching News Plaintiffs' Asserted Works and was acquired by processing raw HTML content scraped from the internet that OpenAI obtained from Common Crawl. *See* Ex. 765 [OPCO_MDL_004598867 - OPCO_MDL_004599087]; Wenger Supplemental Rep. at 24-25; *see also* Delorey Opening Report ¶¶ 218-219.	892. **Disputed in Part – Expert Opinion Subject to Rule 702.** OpenAI responds as follows: Undisputed that the `cc-news` midtraining dataset was [redacted] and that the data used to create `cc-news` was obtained from Common Crawl, which crawls public websites that have not indicated that they opt-out. However, disputed as to the use of the term "matching." Plaintiffs' experts have not proposed a reliable, accurate method of detecting OpenAI training data records "matching" or corresponding to Plaintiffs' Asserted Works at issue in this Motion. *See, e.g.*, S. Krishnan Opp. Decl. (News) Ex. H (S. Krishnan Rebuttal Rep. (Wenger)) § VI (discussing flaws in Wenger's text-matching methodologies and categories of false positives); News Ex. 1167 (S. Krishnan Rebuttal Rep. (Delorey)) § VI (same for Delorey). Because of these issues, both Wenger and Delorey's analyses are subject to a Rule 702 challenge. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Further, disputed that the cited evidence supports the contention that `cc-news` "contained records matching News Plaintiffs' Asserted Works." Microsoft defers to OpenAI.
893. OpenAI [redacted] on News Plaintiffs content. *See* Facts 1188-89, 1221, *infra*.	893. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI [redacted]. Disputed

	that the cited evidence supports the contention that the identified models were mid-trained "on News Plaintiffs content." Microsoft defers to OpenAI.

3. Use of News Content for Post-Training

i) OpenAI's use of News Content for Post-Training

894. Post-training refers to "the set of adaptions" that OpenAI used to "introduce and expose capabilities to users and developers." Ex. 767 [OPCO_MDL_004585370] at 2; Ex. 483 [Callison-Burch Opening Rep.] ¶ 365 ("By 'post-training efforts,' I am referring to the various post-training techniques described earlier in this report that shape the behavior and/or refine the skills of LLMs after pretraining has concluded[.]").	894. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that the cited evidence supports the statement and accurately quotes from the documents. Further disputed that the identified portions of the cited documents are intended to constitute a comprehensive, universal definition of "post-training." For example, Dr. Callison-Burch states elsewhere in his report (in portions not cited by Plaintiffs) that "post-training" can refer to "a suite of techniques to mold the behavior of [a] pre-trained LLM." News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 214. Microsoft defers to OpenAI.
895. OpenAI often uses the term "fine-tuning" interchangeably with "post-training." Ex. 409 [Perelman 30(b)(6) Depo. Tr.] at 19:7-8.	895. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that the cited testimony supports this statement. To the contrary, Mr. Perelman expressly prefaced his testimony with a caveat that he was not qualified to speak on behalf of the OpenAI post-training team, including regarding specific

	terminology: "I should caveat by saying I'm not an expert on post[-]training. That's not part of the team that I work on." Ex. 254 (A. Perelman 30(b)(6) Tr.) 19:3-5. Microsoft defers to OpenAI.
896. OpenAI also refers to post-training as the stage that "aligns" the model. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 139:5-6.	896. **Undisputed.** Microsoft defers to OpenAI.
897. Specifically, during post-training OpenAI "aligned the model to be a helpful assistant for ChatGPT." Ex. 639 [Koch 30(b)(6) Deposition Tr.] at 60:2-11.	897. **Undisputed.** Microsoft defers to OpenAI.
898. OpenAI's post-training efforts are directed to teaching the model to "follow a broad class of written instructions" about how to behave. Ex. 483 [Callison-Burch Opening Rep.] ¶ 377. OpenAI's post-training included the following activities. *Id.* 	898. **Undisputed.** Microsoft defers to OpenAI.
899. Not all of the models at issue were post-trained. Ex. 641 [Heidecke Feb. 12, 2026 30(b)(6) Depo. Tr.] at 110:21-24	899. **Undisputed.** Microsoft defers to OpenAI.
900. GPT-3, the first model deployed through the API, was not post-trained. Ex. 641 [Heidecke Feb. 12, 2026	900. **Undisputed.** Microsoft defers to OpenAI.

30(b)(6) Depo. Tr.] at 110:21-24 ("My understanding is that the GPT-3 model was not post-trained at all.").	
901. One of OpenAI's post-training objectives was to "prevent the model from talking about its own training." Ex. 769 [OPCO_MDL_001936736] at -736; Ex. 771 [OPCO_MDL_000480333] at -333.	901. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI sought to prevent models from talking about training. Disputed that the cited evidence supports the proposition, as neither document states nor suggests that "prevent[ing] the model from talking about its own training" was a "post-training objective." To the extent the cited documents discuss instructions regarding the source of training data, it is in the context of anti-hallucination efforts--i.e., efforts to prevent models from "saying false things, and causing people alarm." News Ex. 771. Note that Exhibit 771 was never filed publicly or served. Microsoft defers to OpenAI.
902. OpenAI used post-training techniques to teach its models to deny or obscure knowledge about training data sources. Ex. 769 [OPCO_MDL_001936736] at -736; Ex. 771 [OPCO_MDL_000480333] at -333.	902. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI sought to prevent models from talking about training. Disputed that the cited evidence supports the proposition, as neither document states or suggests that OpenAI used "post-training techniques" to "teach its models to deny or obscure knowledge about training data sources." To the extent the cited documents discuss instructions regarding the source of training data, it is the context of anti-hallucination efforts–i.e., efforts to prevent models from "saying false things, and causing people alarm." News Ex. 771. Note that

	Exhibit 771 (OPCO_MDL_000480333) was never filed publicly or served. Microsoft defers to OpenAI.
903. Specifically, OpenAI [redacted] Ex. 769 [OPCO_MDL_001936736] at -736.	903. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI sought [redacted]. Disputed that the cited evidence supports the implication that the quoted instructions were implemented in any particular in-scope model, let alone all of them. Further disputed as to Plaintiffs' implicit characterization that OpenAI intentionally sought to deny or obscure knowledge about training data sources. To the extent the cited documents discuss instructions regarding the source of training data, it is the context of anti-hallucination efforts–i.e., the concern that the unspecified model "[h]allucinates the fact that it has proprietary training data." News Ex. 769 at OPCO_MDL_001936736. Microsoft defers to OpenAI.
904. In a Slack Channel titled #tmp-nyt, in June 2023 OpenAI employees discussed [redacted] Ex. 773 [OPCO_MDL_004857423] at -423; Ex. 775 [OPCO_MDL_004897960]; Ex. 777 [OPCO_MDL_004897963] at -963-965.	904. **Disputed in Part – Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that in June 2023, OpenAI employees in the #tmp-nyt channel discussed [redacted] However, disputed that OpenAI [redacted] The disputed proposition is

	not supported by the cited evidence. Specifically, to the extent the cited documents discuss instructions regarding the source of training data, it is the context of anti-hallucination efforts. *See* News Ex. 777 at OPCO_MDL_004897963 (discussing "false positives"). Microsoft defers to OpenAI.
905. OpenAI also used post-training techniques that focused on copyright refusals. Ex. 483 [Callison-Burch Opening Rep.] ¶ 413.	905. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI employed post-training techniques to train the model to refuse requests for copyrighted material. Disputed that the cited evidence supports the proposition as stated, as record evidence demonstrates that OpenAI used a number of anti-regurgitation techniques, including but not limited to post-training techniques focused on copyright refusals. News Ex. 483 (C. Callison-Burch Opening Rep.) § VII.E. Microsoft defers to OpenAI.
906. Not all of the models at issue were post-trained to refuse to regurgitate copyrighted information. Ex. 483 [Callison-Burch Opening Rep.] ¶ 413. Ex. 635 [Callison-Burch Depo. Tr.] at 146:11-19; 163:18-164:5.	906. **Disputed as stated.** OpenAI responds as follows: Disputed that the cited testimony supports this statement. Further, disputed as contradicted by record evidence. OpenAI's corporate representative testified that models underwent general safety post-training, including supervised fine-tuning (SFT), reinforcement learning through human feedback (RLHF), and rule-based reward modeling (RBRM), designed to teach the model not to engage in undesirable behavior, which would have included regurgitation of copyrighted material. *See* Ex. 213

	(J. Heidecke 30(b)(6) Tr. (Feb 12, 2026)) 25:9-16, 111:13–112:18. *See also* News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 414. Microsoft defers to OpenAI.
907. [redacted] OpenAI launched specific post-training initiatives to teach the model to refuse to regurgitate copyrighted information. Ex. 483 [Callison-Burch Opening Rep.] ¶ 413;	907. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that "[s]tarting with [redacted] OpenAI launched specific post-training initiatives to teach the model to refuse to regurgitate copyrighted information." News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 415. Otherwise, disputed that the cited evidence supports this statement. Microsoft defers to OpenAI.
908. That is, [redacted] Ex. 483 [Callison-Burch Opening Rep.] ¶ 413. Ex. 635 [Callison-Burch Depo. Tr.] at 146:11-19; 163:18-164:5.	908. **Disputed in Part.** OpenAI responds as follows: Undisputed that "[s]tarting with [redacted], OpenAI launched specific post-training initiatives to teach the model to refuse to regurgitate copyrighted information." News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 415. However, disputed that the specific cited evidence supports this statement. *See* News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 413; Gratz Decl. Ex. 35 (C. Callison-Burch Tr. (July 15, 2026)) 146:11–19; Ex. 181 (C. Callison-Burch Tr. (July 15, 2026)) 163:18–164:5. Microsoft defers to OpenAI.

909. Every API model deployed before mid-2023 lacked targeted copyright or regurgitation mitigations. Ex. 641 [Heidecke Feb. 12, 2026 30(b)(6) Depo. Tr.] at 24:15-25:16.	909. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that the cited evidence supports the proposition as stated, as record evidence demonstrates that OpenAI employed many regurgitation mitigations prior to that date. Undisputed that a targeted post-training campaign "with the primary focus on regurgitation and copyright" started in March of 2023. While Mr. Heidecke testified that he was "not aware of any targeted work prior" to March 2023, he did not rule out that such work may have occurred. Further, even if not targeted, post-training techniques would have influenced regurgitation behaviors. For example, training data deduplication would constitute a regurgitation mitigation that would have been implemented before mid-2023 *See* News Ex. 483 (C. Callison-Burch Opening Rep.) ¶¶ 410, 414; Ex. 213 (J. Heidecke 30(b)(6) Tr. (Feb. 12, 2026)) 25:9-16. 111:13-112:18. Microsoft defers to OpenAI.
910. Starting with GPT-4, OpenAI's post-training techniques [redacted]. Ex. 483 [Callison-Burch Opening Rep.] ¶ 415; Ex. 781 [OPCO_MDL_000999211] at -211; *see also* Ex. 779 [OPCO_MDL_002138928] at -928; Ex. 641 [Heidecke Feb. 12, 2026 30(b)(6) Depo. Tr.] at 36:12-37:1 ([redacted]).	910. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that the "[redacted]. Disputed that the cited evidence supports the proposition as stated

	that post-training techniques "[redacted]" Microsoft defers to OpenAI.
911. Pursuant to the copyright policy, through its post-training initiatives, OpenAI's models were permitted to return [redacted] copyrighted content per source in a single response. Ex. 781 [OPCO_MDL_000999211] at -214; Ex. 779 [OPCO_MDL_002138928] at -928; Ex. 635 [Callison-Burch Depo.] at 163:12-17.	911. **Undisputed.** Microsoft defers to OpenAI.
912. The Copyright Policy recognizes that [redacted] as an example of what [redacted] Ex. 781 [OPCO_MDL_000999211] at -211.	912. **Undisputed.** Microsoft defers to OpenAI.
913. The majority of training data for OpenAI's post-training refusal mitigation came from [redacted]. Ex. 641 [Heidecke Feb. 12, 2026 30(b)(6) Depo. Tr.] at 72:10-14.	913. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI's corporate representative, Johannes Heidecke, testified that most of the [redacted] used in the targeted post-training campaign on regurgitation involved [redacted]. Ex. 213 (J. Heidecke 30(b)(6) Tr. (Feb. 12, 2026)) 72:10–14. However, OpenAI disputes that the cited evidence supports the proposition as stated. Pursuant to Fed. R. Evid. 106, OpenAI further responds that it used [redacted] to post-train models to refuse requests for copyrighted content. Mr. Heidecke

	testified that most of the training data was [redacted]. That testimony does not necessarily mean that most of the training data *coming out of* the project was [redacted]. Ex. 213 (J. Heidecke 30(b)(6) Tr. (Feb. 12, 2026)) 72:10–14. Microsoft defers to OpenAI.
914. OpenAI could "not recall including any specific text from news publications" in the training data for the post-training refusal mitigation. Ex. 641 [Heidecke Feb. 12, 2026 30(b)(6) Depo. Tr.] at 72:15-18; *see also id*. at 11:6-11.	914. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that the cited deposition transcript contains the quoted language. However, OpenAI disputes Plaintiffs' characterization of "training data," which takes the testimony out of context. Pursuant to Fed. R. Evid. 106, OpenAI further responds that it used [redacted] to post-train models to refuse requests for copyrighted content. Mr. Heidecke testified that most of the training data was [redacted]. That testimony does not necessarily mean that most of the training data *coming out of* the project was [redacted]. Ex. 213 (J. Heidecke 30(b)(6) Tr. (Feb. 12, 2026)) 72:10–14. Microsoft defers to OpenAI.
915. The only [redacted] Ex. 641 [Heidecke Feb. 12, 2026 30(b)(6) Depo. Tr.] at 73:8-9.	915. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that the cited deposition transcript contains the quoted language.

	However, OpenAI disputes Plaintiffs' characterization of "training data," which takes the testimony out of context. Pursuant to Fed. R. Evid.106, OpenAI further responds that it used [redacted] to post-train models to refuse requests for copyrighted content. Mr. Heidecke testified that most of the training data was [redacted]. That testimony does not necessarily mean that most of the training data *coming out of* the project was [redacted] Ex. 213 (J. Heidecke 30(b)(6) Tr. (Feb. 12, 2026)) 72:10–14. Microsoft defers to OpenAI.
916. OpenAI's post-training copyright policy also permits, and even encourages, the model to summarize copyrighted content in response to user requests. Ex. 781 [OPCO_MDL_000999211] at -214, -217 ([redacted]	916. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that [redacted] as allowable model behaviors. News Ex. 781 at OPCO_MDL_000999214, -217. However, OpenAI disputes that the cited evidence supports Plaintiffs' characterization that [redacted] Pursuant to Fed. R. Evid.106, OpenAI further responds that it is inaccurate to say [redacted]

	[redacted]. News Ex. 781 at OPCO_MDL_000999214, -217. Microsoft defers to OpenAI.
917. OpenAI's browsing post-training datasets include over 22,000 mentions of The New York Times and Daily News Plaintiffs' domains. [Goldstein Opening Rep.] ¶ 178.	917. **Disputed in Part - Not Supported by Cited Evidence; Expert Opinion Subject to Rule 702**. OpenAI responds as follows: Undisputed that Dr. Goldstein offered an opinion that reflects this statement. However, OpenAI disputes that the substance of Dr. Goldstein's opinion is supported by record evidence; indeed, Dr. Goldstein cites no record evidence in his report to support this opinion. Also disputed that the statement "include over 22,000 mentions of The New York Times and Daily News Plaintiffs' domains" is vague, irrelevant, and missing necessary context. The 22,000 figure is meaningless without any corresponding details regarding the size of the dataset or "mentions" of other web-domains against which that figure can be compared. In addition, the mere "mention" of a domain in a post-training dataset is not evidence that a domain was accessed or that content was retrieved from that domain. This number comes from Goldstein's analysis. It is not a neutral fact. The 22,000 figure is also untethered to any comparison of the size of the total dataset. Thus, this appears to conflate "mentions" of Plaintiffs' content to copying of Plaintiffs' content. Further, the evidence cited in support of the disputed proposition is currently subject to a motion to exclude

	expert testimony pursuant to Fed. R. Evid. 702. Defendants have moved to exclude Dr. Goldstein's analyses of output logs. *See* ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 13-18. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
918. OpenAI's browsing post-training datasets are used to teach the model how to use the "browser" tool enabling it to search the internet, retrieve results from webpages, and use webpage content for content in crafting a response to a user. Ex. 712 [OPCO_MDL_004673028] at -028; Ex. 783 [OPCO_MDL_004635019] at -019. *See also* Ex. 410 [Feamster Opening Rep.] ¶¶ 210-11.	918. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 918 are disputed because the cited evidence does not support the statement as written. Specifically, News Ex. 712 at OPCO_MDL_004673028 does not describe any "browsing post-training dataset." Instead, it states at a high level that OpenAI "train(ed) ChatGPT to reach for a text-based browser tool when it wants to pull in information from the web." News Ex. 783 likewise does not support the statement. It is a proposal to add new roles ("Platform" and "Developer" system messages) to the model's instruction hierarchy and mentions "post-training datasets" only to say that the main challenge will be migrating those datasets to the new roles. It says nothing about what the browsing post-training datasets are used to "teach." The browser-related text that follows in News Ex.

	783 is an example system message, not a description of a post-training dataset. Finally, pursuant to Fed. R. Evid. 106, Paragraphs 210 and 211 of Dr. Feamster's Opening Report do not mention browsing post-training datasets. Those paragraphs describe the system prompts for the three versions of the Browser tool. OpenAI further disputes the statement to the extent it suggests that the browsing post-training datasets are used only, or mainly, to teach the model how to "search the internet, retrieve results from webpages, and use webpage content" in its responses. That characterization is incomplete and misleading. For example, post-training is also used to limit regurgitation. News Ex. 410 (N. Feamster Opening Rep.) ¶ 239; Fed. R. Evid. 106. Microsoft defers to OpenAI.
919. These browsing post-training datasets consist of [redacted] *See* [Goldstein Opening Rep.] ¶ 176.	919. **Disputed in Part - Not Supported by Cited Evidence.** Defendants respond as follows: Undisputed that Dr. Goldstein offered an opinion that reflects this statement. Also undisputed that the "browsing post-training datasets" contain [redacted] as described in Paragraphs 144–57 of OpenAI's Statement of Undisputed Material Facts in Support of Its Motions for Summary Judgement [ECF 1735]. However, OpenAI disputes that the substance of Dr. Goldstein's opinion is supported by record evidence. Dr.

	Goldstein cites no record evidence in his report to support this opinion. OpenAI further disputes this statement to the extent it suggests that the browsing post-training datasets contain [redacted]. These characterizations, and the inferences drawn from them, are not supported by the record. Microsoft defers to OpenAI.

ii) Microsoft's use of News Content for Post-Training

920. Microsoft's Copilot product used a fine-tuned version of GPT-3.5-Turbo called Deucalion from February 2024 through October 2024. Ex. 1140 [Microsoft's Sept. 12, 2025 Second Supplemental Responses and Objections to DNP's Second Set of Interrogatories] at 8 (First Supplemental Answer to Interrogatory No. 3).	920. **Disputed.** Microsoft responds as follows: this fact is disputed because it misstates the record; to create the Deucalion model, Microsoft participated in post-training on GPT 3.5 Turbo using a synthetic dataset derived from Bing Chat user conversations; the purpose of creating the Deucalion model was to improve the performance of GPT 3.5 Turbo in following instructions and accurately answering user questions as part of the Copilot product. *See* ECF 1723 (J. Lafferty Decl.) ¶¶ 339-40; Hurst Decl. Ex. 85 (M. Parakhin 30(b)(1) Tr.) 204:5-22, 206:13-207:22, 15:18-21, 16:1-5, 18:7-12, 22:25-23:8, 23:21-25:6, 26:12-27:4; Hurst Decl. Ex. 199 (MSFT_AI_MDL_002644093); Hurst Decl. Ex. 113 (X. Song 30(b)(6) Tr.) 46:5–47:13, 53:8–19. OpenAI joins.

921. [redacted] Ex. 1140 [Microsoft's Supplemental Responses to Daily News Plaintiffs' Interrogatories Set No. 2]; Ex. 643 [Song 30(b)(6) Deposition Tr.] at 45:3-9.	921. **Disputed.** Microsoft responds as follows: this fact is disputed because it misstates the record; to create the Deucalion model, Microsoft participated in post-training on GPT 3.5 Turbo using a synthetic dataset derived from Bing Chat user conversations; the purpose of creating the Deucalion model was to improve the performance of GPT 3.5 Turbo in following instructions and accurately answering user questions as part of the Copilot product. *See* ECF 1723 (J. Lafferty Decl.) ¶¶ 339-40; Hurst Decl. Ex. 85 (M. Parakhin 30(b)(6) Tr.) 204:5-22, 206:13-207:22, 15:18-21, 16:1-5, 18:7-12, 22:25-23:8, 23:21-25:6, 26:12-27:4; Hurst Decl. Ex. 199 (MSFT_AI_MDL_002644093); Hurst Decl. Ex. 113 (X. Song 30(b)(6) Tr.) 46:5–47:13, 53:8–19. OpenAI joins.
922. The post-training datasets used to train Deucalion included grounding data sourced from the Bing API. *See* Goldstein Opening Rep. ¶ 181.	922. **Disputed.** Microsoft responds as follows: this fact is disputed because it misstates the record; to create the Deucalion model, Microsoft participated in post-training on GPT 3.5 Turbo using a synthetic dataset derived from Bing Chat user conversations; the purpose of creating the Deucalion model was to improve the performance of GPT 3.5 Turbo in following instructions and accurately answering user questions as part of the Copilot product. *See* ECF 1723 (J. Lafferty Decl.) ¶¶ 339-40; Hurst Decl. Ex. 85 (M. Parakhin 30(b)(6) Tr.) 204:5-22, 206:13-207:22, 15:18-21, 16:1-5, 18:7-12, 22:25-23:8, 23:21-25:6, 26:12-27:4; Hurst Decl. Ex. 199 (MSFT_AI_MDL_002644093);

	Hurst Decl. Ex. 113 (X. Song 30(b)(6) Tr.) 46:5–47:13, 53:8–19. OpenAI joins.
923. The Bing grounding data included in the Deucalion fine-tuning datasets included references to nytimes.com and denverpost.com and snippets from their articles. *See* Goldstein Opening Rep. ¶ 181.	923. **Disputed.** Microsoft responds as follows: this fact is disputed because it misstates the record; The New York Times's experts [redacted] And, News Plaintiffs' expert Dr. Goldstein states [redacted], but he did not [redacted]. *See* ECF 1723 (J. Lafferty Decl.) ¶¶ 337-39; Hurst Decl. Ex. 231 (T. Goldstein Opening Rep.) ¶ 181. Additionally, the Song deposition discusses the Deucalion grounding data, clarifying that this data was only used to teach the model how to respond to tasks, and any news content is incidental to the process. *See* A. Blanco Opp. Decl. Ex. 279 (X. Song 30(b)(6) Tr.) 69:10-70:19. OpenAI joins.
924. Microsoft copied the nytimes.com and denverpost.com webpage content included in the Deucalion training datasets from the Bing Index. Ex. 643 [Song 30(b)(6) Deposition Tr.] at 42:10-44:14.	924. **Disputed.** Microsoft responds as follows: this fact is disputed because it misstates the record; The New York Times's experts [redacted]

	[redacted]. And, News Plaintiffs' expert Dr. Goldstein states [redacted]. *See* ECF 1723 (J. Lafferty Decl.) ¶¶ 337-39; A. Hurst Decl. Ex. 231 (T. Goldstein Opening Rep.) ¶ 181. Additionally, Mr. Song clarified that any news content would be "incidental to what the user query was about." A. Blanco Opp. Decl. Ex. 279 (X. Song 30(b)(6) Tr.) 44:8-14. OpenAI joins.

4. "Memorization" of News Content During Training

925. "Memorization" refers to instances where the model retains specific text from its training data in its parameters as a state of the trained model. Etzioni Rebuttal Rep. ¶ 53; Wenger Opening Rep. at 67; see also Ex. 675 [Callison-Burch Reply Rep.] ¶¶ 34 & n.33 (identifying this as "'verbatim' or 'near-verbatim' memorization" and quoting the Etzioni definition), 52 ("we can agree that some memorization occurs"), 56 ("I did not and do not contest that memorization occurs in LLMs."); Goldstein Opening Rep. ¶ 297 ("'Memorization' 'refers to instances where an LLM's parameters encode copies of its training data.").	925. **Disputed - Expert Opinion Subject to Rule 702.** OpenAI responds as follows: Disputed to the extent Plaintiffs purport to define "memorization" as the retention of specific training text in a model's parameters. As Dr. Callison-Burch explains, memorization is an "ambiguous concept" used in different ways in the literature, and Plaintiffs' proposed definition concerns only one form — "verbatim" or "near-verbatim" memorization. News Ex. 675 (C. Callison-Burch Reply Rep.) ¶ 34 & n.33. Pursuant to Fed. R. Evid.106, Dr. Callison-Burch further explains that memorization can refer to a model internalizing the properties of a particular sequence of text, such that the

	model assigns a high probability to the next word in that sequence, rather than retaining a literal copy of the text in its parameters. *Id.* ¶¶ 34–35. The cited research also distinguishes among multiple forms of memorization, including the internalization of general facts, ideas, or alignment goals. Accordingly, Plaintiffs' characterization improperly treats one particular conception of memorization as a comprehensive definition of the term. The fact stated in Paragraph 925 is further disputed because the evidence cited in support is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Defendants have moved to exclude certain opinions of Dr. Thomas Goldstein ("copies," CMI-removal, Copilot log, and Microsoft-role opinions); Dr. Emily Wenger (text-matching, "regurgitation," CMI-removal, and Copilot log extrapolation opinions); Dr. Oren Etzioni (value-of-news and Copilot substitution opinions). *See* ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 9-22, 25-28. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
926. "Regurgitation" is closely related to memorization and refers to model output behavior where an LLM generates a text sequence that is verbatim or near-verbatim to a text sequence contained in a model's training corpus. Ex. 675 [Callison-Burch Reply Rep.] ¶ 43;	926. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that the cited evidence contains the quoted statements. However, disputed to the extent Plaintiffs characterize regurgitation as "closely related" to memorization or suggest that

Goldstein Opening Rep. ¶ 297 ("Memorization is related to another concept known as 'regurgitation,' in which an LLM outputs text containing a verbatim reproduction of a part of its training data. Put simply, regurgitation is an instance in which memorization becomes directly observable in model.").	memorization results in regurgitation. Pursuant to Fed. R. Evid. 106, as Dr. Callison-Burch explains, Plaintiffs conflate the concepts of memorization and regurgitation and overstate the extent to which memorized content can be elicited. See News Ex. 675 (C. Callison-Burch Reply Rep.) ¶¶ 43–46. Indeed, even where models memorize repeated sequences, such memorization may be difficult to discover and extracting that content may require qualitatively different attack strategies. *See* A. Blanco Opp. Decl. Ex. 136 (N. Carlini et al., Quantifying Memorization Across Neural Language Models (ICLR 2023)) at 5. Further, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
927. Calibrating the model according to the next token prediction training objective leads to memorization and regurgitation of training data. Ex. 645 [Weng Depo. Tr.] at 40:18-41:5; Ex. 714 [OPCO_MDL_001900461] at -468 ("It's worth noting here that models want to regurgitate, so we're inherently fighting against the training objective"); Ex. 718 [OPCO_MDL_001644364] at 365 (OpenAI's VP of Research, Jerry Tworek: "We train our networks to memorize the training data — that's their objective.").	927. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that the next-token prediction training objective necessarily leads to memorization and regurgitation of training data. Plaintiffs' characterization conflates memorization and regurgitation, which are distinct concepts, and takes the cited statements out of context. Pursuant to Fed. R. Evid. 106, as Dr. Callison-Burch explains, even where a model memorizes training data, it does not necessarily regurgitate that data in its outputs, including where the model has undergone anti-regurgitation training. *See* News Ex. 675 (C. Callison-Burch Reply Rep.) ¶ 55. Moreover, Mr. Tworek's statement that "[w]e train our networks to memorize the training data"

	concerned a model trained on the same data multiple times, rather than establishing that next-token prediction training generally results in memorization or regurgitation. *See* News Ex. 718 at OPCO_MDL_001644365–366. Both quotes are taken out of context: for example, Mr. Tworek is referring to a model trained on the same data multiple times. Dr. Callison-Burch discusses how Plaintiffs' claims that the training objectives of next-token prediction inevitably lead to regurgitation and memorization are misleading, as memorization and regurgitation are separate concepts. He also points out that a model that memorizes data does not necessarily regurgitate it (if, for example, it has undergone anti-regurgitation training). News Ex. 675 (C. Callison-Burch Reply Rep.) ¶ 55. The facts set forth in Paragraph 927 are disputed because the cited evidence is incorrect. Microsoft defers to OpenAI.
928. "Autoregressive" refers to generating text one token at a time, where each token depends on the words coming before it. Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 138-39.	928. **Undisputed.** Microsoft defers to OpenAI.
929. Memorization and regurgitation are related to the next-token prediction objective because LLMs are trained to autoregressively predict the next token. Ex. 645 [Lilian Weng Depo. Tr.] at 40:25-41:5.	929. **Disputed.** OpenAI responds as follows: Disputed to the extent Plaintiffs' characterization suggests that autoregressive next-token prediction necessarily causes memorization and regurgitation of training data. The cited testimony does not establish such a causal relationship. Pursuant to Fed. R. Evid. 106, as Dr. Callison-Burch explains, memorization and regurgitation are

	distinct concepts, and a model that memorizes training data does not necessarily regurgitate that data in its outputs, including where the model has undergone anti-regurgitation training. *See* News Ex. 675 (C. Callison-Burch Reply Rep.) ¶ 55. Moreover, Ms. Weng testified that [redacted] a qualification omitted from Plaintiffs' characterization. News Ex. 645 (L. Weng 30(b)(1) Tr.) 40:25–41:5. Plaintiffs' assertion also fails to distinguish between a model learning facts or commonly occurring phrases, which may be beneficial to model performance, and undesirable memorization of training data. Pursuant to Fed. R. Evid.106, *see* News Ex. 675 (C. Callison-Burch Reply Rep.) ¶¶ 48, 51–52, 55. Microsoft defers to OpenAI.
930. "Verbatim memorization" happens to some degree in LLMs. Ex. 675 [Callison-Burch Reply Rep.] ¶¶ 51-52 ("verbatim memorization" happens to some degree and that "some memorization occurs.").	930. **Disputed in Part.** OpenAI responds as follows: Undisputed that the models at issue can memorize certain facts or sequences of text from their training data. Disputed to the extent News Plaintiffs' characterization fails to distinguish between beneficial memorization of facts, common phrases, and idioms -- and undesirable memorization of training data. As Dr. Callison-Burch explains, learning and retaining facts and commonly occurring sequences of text can be beneficial to model performance and reflect generalization rather than improper memorization. *See* News Ex. 675 (C. Callison-Burch Reply Rep.) ¶¶ 48, 51–52; *see also* ECF 1819-2 (T. Goldstein Rebuttal Rep.) ¶ 62; A. Blanco Opp. Decl. Ex. 133 (M.

	Mitchell Rebuttal Rep.) ¶ 25. Accordingly, the mere ability of a model to memorize facts or text sequences does not establish that it memorizes protectable expression in a manner that results in regurgitation. Microsoft defers to OpenAI.
931. The models at issue can memorize facts or sequences of text from the training corpus. Ex. 675 [Callison-Burch Reply Rep.] ¶¶ 48, 51-52.	931. **Disputed in Part.** OpenAI responds as follows: Undisputed that the models at issue can memorize certain facts or sequences of text from their training data. Disputed to the extent News Plaintiffs' characterization fails to distinguish between beneficial memorization of facts, common phrases, and idioms -- and undesirable memorization of training data. As Dr. Callison-Burch explains, learning and retaining facts and commonly occurring sequences of text can be beneficial to model performance and reflect generalization rather than improper memorization. *See* News Ex. 675 (C. Callison-Burch Reply Rep.) ¶¶ 48, 51–52; *see also* ECF 1819-2 (T. Goldstein Rebuttal Rep.) ¶ 62; A. Blanco Opp. Decl. Ex. 133 (M. Mitchell Rebuttal Rep.) ¶ 25. Accordingly, the mere ability of a model to memorize facts or text sequences does not establish that it memorizes protectable expression in a manner that results in regurgitation. Microsoft defers to OpenAI.
932. One factor that influences the extent to which LLM's memorize their training data is the number of	932. **Disputed in Part.** OpenAI responds as follows: Disputed to the extent News Plaintiffs characterize Dr. Callison-Burch's opinion as stating that the number of

times the model saw certain data during training. Ex. 483 [Callison-Burch Opening Rep.] ¶ 411.	times a model encounters training data influences the extent to which the model "memorizes" that data. The cited paragraph instead discusses overfitting and the risk that a model may "inadvertently generate text that matches text in the training data," rather than memorization. *See* News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 411. As Dr. Callison-Burch further explains, memorization, overfitting, and regurgitation are distinct concepts that Plaintiffs improperly conflate. *See* ECF 1715 (C. Callison-Burch Decl.) ¶ 114; News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 411. Microsoft defers to OpenAI.
933. The ordering and amount of training data shown to the model during training are design decisions made by OpenAI. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 178:19-179:14).	933. **Disputed in Part.** OpenAI responds as follows: Undisputed that OpenAI makes design decisions concerning the ordering and amount of data presented to models during pretraining. Disputed to the extent News Plaintiffs' characterization extends Mr. Ryder's testimony to all stages of model training. Mr. Ryder was designated as a Rule 30(b)(6) witness on pretraining topics, and his testimony concerning the ordering and amount of training data shown to models was limited to pretraining. *See* A. Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr.) at 178:19–179:14. Microsoft defers to OpenAI.
934. Each "pass" over training data is called an "epoch." Ex. 647 [Parakhin Depo. Tr.] at 60:18-61:11.	934. **Undisputed.**

935. During training, OpenAI performed multiple epochs on certain high-quality subsets of its pertaining data. Ex. 680 [OPCO_MDL_000000439] at -446-47].	935. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: OpenAI trained for multiple epochs on certain subsets of its training mix that it believed, at the time of training GPT-3, to be higher quality than CommonCrawl data. However, its understanding of "quality" of training data has evolved since GPT-3. *See* A. Hurst Decl. Ex. 104 (N. Ryder (30)(b)(6) Tr.) at 32:24-33:6. Microsoft defers to OpenAI.
936. GPT-4 was trained using multiple epochs over [redacted] with the number of epochs varying [redacted]. Ex. 678 [OPCO_MDL_004806204] at -205.	936. **Undisputed.** Microsoft defers to OpenAI.
937. In 2022, Alex Paino advised his colleagues "Looking forward, though, I think we should assume future big models will be able to memorize everything in the train corpus, even if it only appears once per epoch." Ex. 598 [OPCO_MDL_001697816] at –816.	937. **Disputed in Part.** OpenAI responds as follows: Undisputed that Ex. 598 contains the quoted language. However, disputed to the extent News Plaintiffs characterize Mr. Paino's statement as a conclusion about the memorization capabilities of future models, rather than a speculative prediction. Mr. Paino expressly qualified his statement as "[b]ased on some newer work on limited data scaling laws" and stated that he "think[s]" OpenAI "should assume" future large models would be capable of memorizing their entire training corpus. News Ex. 598 at OPCO_MDL_001697816. News Plaintiffs' characterization omits this qualification and fails to distinguish Mr. Paino's prediction about hypothetical future models from evidence concerning the actual memorization capabilities of the models at issue. *See* Fed. R. Evid. 106.

	Microsoft defers to OpenAI.
938. OpenAI knew the models at issue had memorized text from the training corpus. See Ex. 684 [OPCO_MDL_004672513] ("The model will readily produce excerpts from works under copyright.") at 3; Ex. 716 [OPCO_MDL_000479579] at -83; Ex. 720 [OPCO_MDL_001939120] at -124-26; Ex. 728 [OPCO_MDL_001639231] at -231.	938. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed to the extent News Plaintiffs characterize the cited evidence as establishing that OpenAI knew all of the models at issue had memorized text from their training corpora. News Plaintiffs' assertion conflates memorization with distinct concepts such as overfitting, regurgitation, and reproduction of training data. *See* Ex. 250 (A. Paino 30(b)(6) Tr.) at 62:2–63:9 (discussing distinctions among these concepts and noting that "regurgitation" has multiple definitions); *id.* at 68:15–25 (same). For example, the statement that "[t]he model will readily produce excerpts from works under copyright" concerns model output behavior, not necessarily whether or how the model memorized text during training. *See* News Ex. 684 at slide 3. Further disputed to the extent Plaintiffs generalize statements concerning particular models or observed behaviors to all models at issue without establishing that the same behavior occurred in each model. Microsoft defers to OpenAI.
939. Microsoft knew the models at issue had memorized text from the training corpus. Ex. 722 [MSFT_AI_MDL_002836983] at -983; Ex. 724 [MSFT_AI_MDL_000665166]; Ex. 1268 [MSFT_AI_MDL_000665167] (Templeton Ex. 31) at -167, -169.	939. **Disputed – Immaterial.** Microsoft responds that the facts set forth in Paragraph 939 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Microsoft's alleged "knowledge" is not an element of any claim or defense at issue in Plaintiffs' summary judgment motion.

	Further responding, Microsoft states that Hecht's statements in Ex. 724 cannot be attributed to Microsoft. Hecht is a research academic employed at Microsoft to bring "divergent," "asymmetric, futuristic, and academic points of view"; he does not set Microsoft's core priorities for Microsoft or speak for Microsoft as to on "his theoretical views about AI's potential effect on content creators"; Microsoft leadership "does not share all of [his] views"; and he is a "kind of intellectual provocateur." ECF 1707 (J. Usdan Decl.) ¶ 39; Ex. 285 (J. Usdan 30(b)(6) Tr.) at 173:22-174:9. Hecht has himself disclaimed any authority to speak for Microsoft, testifying that his views on these subjects are "my view, not Microsoft's view," that "[he] has no idea if [they are shared] by the board of Microsoft," and that "[he] can't speak for the company on this stuff." Ex. 212 (B. Hecht 30(b)(1) Tr.) at 319:18–20, 320:10–12, 321:4–5. He further described his views as "hypotheses," cautioned that "the jury is still out on a lot of this stuff," and acknowledged that he is "not a lawyer." Ex. 212 (B. Hecht 30(b)(1) Tr.) 59:11–60:8, 165:9–15, 166:10–11. His personal views therefore are not statements of Microsoft and do not reflect Microsoft's knowledge, policies, or conduct. OpenAI joins Microsoft's response.
940. In August 2023, [redacted]	940. **Disputed in Part.** Microsoft does not dispute that the exhibit contains the quoted language. But Microsoft does dispute the notion in Paragraph 940 that Copilot outputs copyrighted works on command. Models can at

[redacted]. Ex. 726 [MSFT_AI_MDL_001815252] at -252-53.	times generate text that was part of their training corpus, a known phenomenon called regurgitation. However, instances of regurgitation in a properly trained model are extremely rare. *See* ECF 1723 (J. Lafferty Decl.) ¶ 35; ECF 1707 (J. Usdan Decl.) ¶¶ 19-20. OpenAI joins Microsoft's response.
941. The CTO of Microsoft Azure, Mark Russinovich, testified that he [redacted] Ex. 649 [Mark Russinovich Depo. Tr.] at 104:16-105:2.	941. **Disputed in Part.** Microsoft responds as follows: Microsoft does not dispute that Mr. Russinovich made those statements in Paragraph 941 in his deposition. But Microsoft does dispute the notion in Paragraph 941 that Copilot outputs copyrighted works on command. Models can at times generate text that was part of their training corpus, a known phenomenon called regurgitation. However, instances of regurgitation in a properly trained model are extremely rare. *See* ECF 1723 (J. Lafferty Decl.) ¶ 35; ECF 1707 (J. Usdan Decl.) ¶¶ 19-20. OpenAI joins Microsoft's response.
942. In November 2019 OpenAI expressed concerns abouts its models "accidentally regenerating copyrighted works." Ex. 716 [OPCO_MDL_000479579] at -83.	942. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 716 identifies "accidentally regenerating copyrighted works" as a potential issue in the context of developing deduplication tools. Disputed to the extent News Plaintiffs characterize this statement as reflecting a broader concern by OpenAI that its models were reproducing copyrighted works. The cited document, titled "Algorithms Data Needs Themes," identifies technical challenges and

	potential solutions for future development, including tools to detect whether generative models produce items already present in their training datasets. News Ex. 716 at OPCO_MDL_000479583. Microsoft defers to OpenAI.
943. Open AI understood in 2020 that even showing the model data once during training "causes some degree of memorization." Ex. 720 [OPCO_MDL_001939120] at -124-26.	943. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 720 contains a statement by Dario Amodei that "even seeing the data once causes some degree of memorization." Disputed to the extent News Plaintiffs characterize this statement as establishing that a single exposure to training data necessarily results in memorization and reproduction. Amodei qualified his statement by explaining that data encountered only once during training exhibited a relatively small reduction in entropy and that this "shouldn't be enough to generate something at all similar to what was trained on." News Ex. 720 at OPCO_MDL_001939126. Amodei further explained that although a model may initially memorize data following a gradient update, that information may be partially forgotten over time, making it unlikely that arbitrary training data would remain fully memorized. *Id.* at OPCO_MDL_001939125. Microsoft defers to OpenAI.
944. In 2020, OpenAI recognized that memorization "relates to privacy and copyright issues [it	944. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that

was] likely to encounter in commercialization eventually." Ex. 720 [OPCO_MDL_001939120] at -125.	News Ex. 720 contains the quoted language. Disputed to the extent News Plaintiffs characterize the statement as reflecting a broader conclusion by OpenAI that memorization necessarily gives rise to privacy or copyright issues. The cited statement was made by Dario Amodei in the context of a discussion concerning the potential memorization of HackerRank problems and his desire for a mechanism to "forget training data." News Ex. 720 at OPCO_MDL_001939125. Amodei's statement concerned potential issues that might arise in future commercialization. Microsoft defers to OpenAI.
945. OpenAI understood in 2020 that its API "might output existing content verbatim." Ex. 784 [OPCO_MDL_001812710] at -710.	945. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 784 contains the quoted statement. Disputed to the extent News Plaintiffs characterize that Slack discussion as establishing that OpenAI generally understood its API to reproduce content verbatim. Ex. 784 reflects a customer inquiry about a potential risk, not a determination that the API produced such outputs. The discussion expressly qualified that risk, explaining that it was more likely where prompts closely resembled pretraining data and less likely where they did not. *See* News Ex. 784. Microsoft defers to OpenAI.
946. In 2021, OpenAI considered the prevention of memorization as important "for fair use [compliance] and	946. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 787 contains the quoted language. Disputed to the

minimizing copyright violations in model output." Ex. 787 [OPCO_MDL_000207987] at -990.	extent News Plaintiffs characterize this statement as reflecting a general conclusion by OpenAI. The quoted statement was made by Alec Radford in the context of proposing an additional technical measure to reduce memorization of duplicated training data, rather than asserting that memorization itself constituted copyright infringement. News Ex. 787 at OPCO_MDL_000207990–991. Mr. Radford [redacted]. *Id.* Microsoft defers to OpenAI.
947. In a document created in October 2021 by Alex Paino, Paino notes that "[OpenAI] and external parties, have noticed a lot of instances where GPT-3 or our Codex models will regurgitate training examples verbatim." Ex. 789 [OPCO_MDL_001938504] at -504.	947. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 789 contains the quoted statement. Disputed as unsupported by the cited evidence. Disputed to the extent News Plaintiffs characterize Mr. Paino's statement as establishing that GPT-3 and Codex models routinely regurgitated training examples verbatim. The cited document, titled "Preventing Regurgitation," discusses observed instances of verbatim reproduction in developing technical measures to mitigate it. News Ex. 789 at OPCO_MDL_001938504. The document further explains that memorization depends on factors including the frequency particular training examples appear and describes experiments showing that certain models did not reproduce

	examples even after repeated exposure. *Id.* at OPCO_MDL_001938519–521, -528. Mr. Paino also describes a masking approach that appeared effective at reducing regurgitation without materially affecting model performance. *Id.* at OPCO_MDL_001938504–505, -507. Microsoft defers to OpenAI.
948. By March 2022, "[r]egurgitation [had] come up repeatedly as a problem" for OpenAI, and OpenAI recognized that it "should be thinking about [regurgitation] for every large model[.]" Ex. 728 [OPCO_MDL_001639231] at –31.	948. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 728 contains the quoted language. Disputed as unsupported by the cited evidence. Disputed to the extent News Plaintiffs characterize the statement as establishing that regurgitation was an ongoing issue across OpenAI's large models. The cited statement was made in a discussion proposing greater awareness of anti-regurgitation techniques, including a joint technical presentation and memorandum. It specifically noted that regurgitation had been addressed for GPT and DALL-E and suggested that OpenAI consider similar measures for future large models intended for release. *See* News Ex. 728. Microsoft defers to OpenAI.
949. In June 2022, OpenAI employees were concerned that GPT-4 "will have memorized a ton of data and therefore will be insanely good at regurgitation." Ex. 730 [OPCO_MDL_001641776] at -776.	949. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that News Ex. 730 contains the quoted language. However, disputed to the extent Plaintiffs characterize the statement as reflecting a conclusion that GPT-4 would memorize substantial amounts of training data and regurgitate that

	content. The cited statement was made by Ilya Sutskever, who described a "possible concern" raised by Alec Radford and expressly solicited others' views, asking "what are our official thoughts on this?" News Ex. 730 at OPCO_MDL_001641776. The ensuing discussion acknowledged existing regurgitation-prevention measures targeting frequently occurring strings. *Id.* at OPCO_MDL_001641776776–77. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
950. A September 2022 report titled "Davinci-3 Preliminary RAI Evaluation: Interim Report" reported that "DV3 seems to have greater memorization capabilities, resulting in higher risk of exposing sensitive or confidential information from its training data" and "DV3 shows a greater capacity for raw training data memorization than its predecessor model." Ex. 734 [OPCO_MDL_001710573] at -95.	950. **Disputed in Part - Not Supported by Cited Evidence**. Defendants respond as follows: Undisputed that News Ex. 734 contains the quoted statements. Disputed to the extent Plaintiffs characterize these statements as established findings concerning DV3's memorization capabilities or the actual exposure of sensitive or confidential training data. The cited document is a preliminary evaluation conducted by a Microsoft team in collaboration with OpenAI's Responsible AI team, which expressly cautions that its findings "are not broad or statistically significant but are preliminary and must be treated as such." News Ex. 734 at OPCO_MDL_001710594. The evaluation of sequence memorization was based on a limited set of prompts drawn from Wikipedia articles concerning anarchism, rather than a comprehensive assessment of DV3's training data or memorization capabilities. *Id.* at OPCO_MDL_001710594–595. Moreover, the report

	characterizes the potential exposure of sensitive information as a risk, not an actual instance of it. *Id.* at OPCO_MDL_001710595. Therefore, the disputed proposition is not supported by the cited evidence.
951. In October 2022, OpenAI employee Peter Welinder (then a member of the Applied group, focused on product design) described a meeting during which "bob [McGrew, Head of Research for OpenAI], me, and che [Chang, OpenAI's General Counsel] aligned on the technical offsite that we're fine with the regurgitation risk, so should not be a release blocker." Ex. 736 [OPCO_MDL_004608715] at -15; Ex. 527 [Welinder Nov. 13, 2025 Depo Tr.] at 33:22-34:20, 41:4-15.	951. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that News Ex. 736 contains the quoted language. Disputed to the extent Plaintiffs characterize the statement as establishing that OpenAI generally accepted regurgitation risks or determined that such risks did not warrant mitigation. The cited statement in an October 2022 Slack discussion concerned model release planning and that regurgitation risk should not prevent a particular release, rather than a determination that regurgitation posed no concern or that OpenAI would not continue to address it. News Ex. 736 at OPCO_MDL_004608715. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
952. Leading up to the launch of ChatGPT, on November 2, 2022 OpenAI recognized that the model could regurgitate its training data but reaffirmed it had "aligned prior to training that regurgitation risk was low enough that commercializing would be fine." Ex. 738 [OPCO_MDL_004617667] at -670.	952. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that News Ex. 738 contains the quoted language. Disputed to the extent Plaintiffs characterize this statement as establishing that OpenAI accepted regurgitation risk or determined that regurgitation was not a concern. The cited statement was from a November 2022 Slack discussion concerning release planning for a particular version of GPT-V, a model not at issue in this case. News Ex. 738 at

	OPCO_MDL_004617667. The cited language does not support Plaintiffs' inference that regurgitation was an accepted system-level risk that OpenAI would not address. Microsoft defers to OpenAI.
953. Prior to the launch of ChatGPT, OpenAI employees were concerned that "if [ChatGPT] gets more popular than playground, regurg issues with PII/copyright [would] get publicity." Ex. 740 [OPCO_MDL_004809431] at –437.	953. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 740 contains part of the quoted statement. However, the proposition set forth in Paragraph 953 are not supported by the cited evidence. News Ex. 740 does not discuss playground. Also, disputed to the extent Plaintiffs characterize this statement as establishing that OpenAI employees knew ChatGPT would regurgitate material. The cited document describes a hypothetical pre-launch concern, alongside other potential risks, and expressly identifies certain outputs as OpenAI concerns. News Ex. 740. Ex. 740 explains that, following ChatGPT's launch, OpenAI devoted substantial efforts to preventing regurgitation, including mitigation measures at the pretraining, post-training, and inference stages. *Id.* at OPCO_MDL_004809436–437. Microsoft defers to OpenAI.
954. OpenAI recognized that "Language models are pretty good at memorizing docs — especially if we train for many epochs." Ex. 742 [OPCO_MDL_001876595] at -596.	954. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 742 contains the quoted statement. Disputed to the extent Plaintiffs characterize this statement as establishing that OpenAI's models generally memorized entire

	documents during ordinary training. The quoted statement was made by an individual OpenAI employee in the context of proposing a potential technical solution for handling requests to remove personally identifiable information and copyrighted content from model outputs. That employee's observation that language models are particularly effective at memorizing documents when trained for many epochs concerned that context, rather than general memorization capabilities of production models during training. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
955. OpenAI acknowledged that around the time ChatGPT launched it was aware that models could memorize and regurgitate data. Ex. 651 [Monaco April 8 30(b)(6) Deposition Tr.] at 212:2-5.	955. **Disputed in Part.** OpenAI responds as follows: Undisputed that Monaco testified that, around the time ChatGPT launched, "it was general knowledge in the field, and we were aware that models can memorize and regurgitate data." Ex. 240 (J. Monaco 30(b)(6) Tr. (Apr. 8, 2026)) 212:2–5. Disputed to the extent that Plaintiffs' mischaracterized Mr. Monaco's testimony. Pursuant to Fed. R. Evid. 106, Mr. Monaco's immediately following statement was: "And we wanted to prevent that." *Id.* Microsoft defers to OpenAI.
956. In January 2023 Microsoft discovered [redacted]	956. **Disputed in Part.** Microsoft responds as follows: Microsoft does not dispute that the exhibit contains the quoted language. But Microsoft does dispute the notion in Paragraph 956 that Copilot outputs copyrighted works on

[redacted] Ex. 722 [MSFT_AI_MDL_002836983] at -983.	command. Models can at times generate text that was part of their training corpus, a known phenomenon called regurgitation. However, instances of regurgitation in a properly trained model are extremely rare. *See* ECF 1723 (J. Lafferty Decl.) ¶ 35; ECF 1707 (J. Usdan Decl.) ¶¶ 19-20. OpenAI joins.
957. In February 2023, Microsoft's Director of Applied Science, Brent Hecht, messaged on Teams to share "a few highlights worth mentioning" from a February 2023 study conducted by Microsoft employees which concerned memorization and regurgitation in LLMs, including GPT-3 and ChatGPT. The highlights reported by Mr. Hecht were "• Improved accuracy is associated with output of memorized text. As we reduce hallucination, it will be important to see if we are just repeating roughly verbatim what was in the pre-training data....• More evidence that hallucination and related issues are far more common than in the public imagination." Ex. 724 [MSFT_AI_MDL_000665166].	957. **Disputed in Part – Immaterial.** Microsoft responds as follows: Microsoft does not dispute that the exhibit contains the quoted language. But Microsoft does dispute the implication that Defendants' Generative AI tools routinely generate regurgitated text. Models can at times generate text that was part of their training corpus, a known phenomenon called regurgitation. However, instances of regurgitation in a properly trained model are extremely rare. A model that has a tendency to regurgitate its training data reflects an unintended side-effect of the training process rather than the intended operation of the model. *See* ECF 1723 (J. Lafferty Decl.)¶ 35; ECF 1707 (J. Usdan Decl.) ¶¶ 19-20. Microsoft also responds that the facts set forth in Paragraph 957 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Hecht's quoted statements at Ex. 724 reflect his preliminary takeaways from a draft manuscript that he had noy yet fully reviewed and whose authors were seeking feedback, Ex. 724 at -166. Hecht's statements cannot be attributed to Microsoft. Hecht is a

	research academic employed at Microsoft to bring "divergent," "asymmetric, futuristic, and academic points of view"; he does not set Microsoft's core priorities for Microsoft or speak for Microsoft as to "his theoretical views about AI's potential effect on content creators"; Microsoft leadership "does not share all of [his] views"; and he is a "kind of intellectual provocateur." ECF 1701 (J. Usdan Decl.) ¶ 39; Ex. 285 (J. Usdan 30(b)(6) Tr. (Feb. 11, 2026)) 173:22-174:9. Hecht has himself disclaimed any authority to speak for Microsoft, testifying that his views on these subjects are "my view, not Microsoft's view," that "[he] has no idea if [they are shared] by the board of Microsoft," and that "[he] can't speak for the company on this stuff." Ex. 212 (B. Hecht 30(b)(1) Tr.) 319:18–20, 320:8–12, 321:4–6. He further described his views as "hypotheses," cautioned that "the jury is still out on a lot of this stuff," and acknowledged that he is "not a lawyer." *Id.* 59:11–60:8, 165:9–15, 166:10–11. His personal views therefore are not statements of Microsoft and do not reflect Microsoft's knowledge, policies, or conduct. OpenAI joins Microsoft's response, except to the extent any portion of it, including any testimony quoted under Fed. R. Evid. 106, could be read to characterize OpenAI, its models, or its conduct. OpenAI disputes any such characterization. Hecht is a Microsoft researcher who testified unequivocally: "I don't speak for Microsoft. I definitely don't speak for OpenAI." *Id.* 100:11–12. His statements are not based on personal knowledge of OpenAI, are not attributable to OpenAI, and are inadmissible hearsay as to OpenAI. *See* Fed.

	R. Civ. P. 56(c)(2), (c)(4); Fed. R. Evid. 602, 801(d)(2)(D), 802. This paragraph cannot be offered to establish any fact concerning OpenAI.
958. The February 2023 study conducted by Microsoft employees evaluating six LLMs outputs for "Discourse, Memorization, and Reliability" found that "[a]bout 78.4% proportion of the outputs presented some degree of memorization, with an average of 79.2% original text" and, among the six LLMs, the "GPT-3 models and ChatGPT [had] the highest incidence of memorization." Ex. 1268 [MSFT_AI_MDL_000665167] (Templeton Ex. 31) at -167, -169.	958. **Disputed in Part.** Microsoft responds as follows: Microsoft does not dispute that the exhibit contains the quoted language. But Microsoft does dispute the characterization in Paragraph 958 that "some degree of memorization, with an average of 79.2% original text" entails widespread regurgitation issues (as opposed to being specific to that sample size). Rather, models can at times generate text that was part of their training corpus, a known phenomenon called regurgitation. However, instances of regurgitation in a properly trained model are extremely rare. *See* ECF 1723 (J. Lafferty Decl.) ¶ 35; ECF 1707 (J. Usdan Decl.) ¶¶ 19-20. OpenAI joins.
959. OpenAI recognized that GPT-3.5, the model initially powering ChatGPT, could "easily reproduce training data verbatim." Ex. 791 [OPCO_MDL_001837518] at -518.	959. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that Ex. 791 contains the quoted statement. Disputed to the extent Plaintiffs imply that this statement concerns the specific GPT-3.5 model version initially deployed in ChatGPT. The cited document does not identify that model version. Moreover, the statement appears in a document concerning efforts to prevent regurgitation. News Ex. 791 at OPCO_MDL_001837518. Therefore, the disputed proposition is not supported by the cited evidence.

	Microsoft defers to OpenAI.
960. Prior to training GPT 4, OpenAI acknowledged that its "models generate verbatim copies of training data" and noted that it "risked being sued if regurgitation [was] not addressed." Ex. 793 [OPCO_MDL_004838627] at -627.	960. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that News Ex. 793 contains the quoted language. However, disputed to the extent Plaintiffs characterize these statements as establishing that OpenAI's models generally reproduced training data verbatim or suggested such reproduction was unlawful. The cited document concerns a proposed technical project to reduce regurgitation in GPT-4 before training. News Ex. 793 at OPCO_MDL_004838627. The quoted statement was made in the context of identifying potential risks associated with model outputs and motivating preventive measures, not suggesting that any reproduction was unlawful. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
961. OpenAI acknowledged it should "make this repugnant regurgitation issue disappear" and noted that it would "have to overcome this issue for GPT-4 training starting." Ex. 795 [OPCO_MDL_000110815] at -815-16.	961. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 795 contains the quoted language. However, disputed to the extent Plaintiffs characterize these statements as conclusions adopted by OpenAI generally, rather than statements made by an individual employee. Specifically, that employee sought volunteers to develop regurgitation mitigation techniques before GPT-4 training. News Ex. 795 at OPCO_MDL_000110815–816. The quoted statement was part of an effort to recruit

	employees to address the issue, rather than a determination about the extent of regurgitation across OpenAI's models. Moreover, the statement that OpenAI would "have to overcome this issue for GPT-4 training starting" concerned prospective efforts to mitigate regurgitation, not evidence that GPT-4 itself exhibited such behavior. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
962. In May 2023, an OpenAI employee, Angela Jiang, flagged that "[o]ne concern is that our models tend to regurgitate, so even the model trained on copyright is a liability[.]" Ex. 797 [OPCO_MDL_001699893] at -893.	962. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that News Ex. 797 contains the quoted language. Further disputed to the extent Plaintiffs characterize Jiang's statement as establishing that OpenAI's models generally regurgitated training data or that training on copyrighted material necessarily created legal liability. The cited statement was made by an individual employee in response to a question concerning the use of certain data for training and identifies regurgitation as "one concern," rather than an established conclusion about the behavior of a model. News Ex. 797 at OPCO_MDL_001699893. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
963. In June 2023 OpenAI recognized that its "current models [] often regurgitate long sequences of copyrighted material." Ex. 779 [OPCO_MDL_002138928] at -928.	963. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that News Ex. 779 contains the quoted language. However, disputed to the extent Plaintiffs characterize this statement

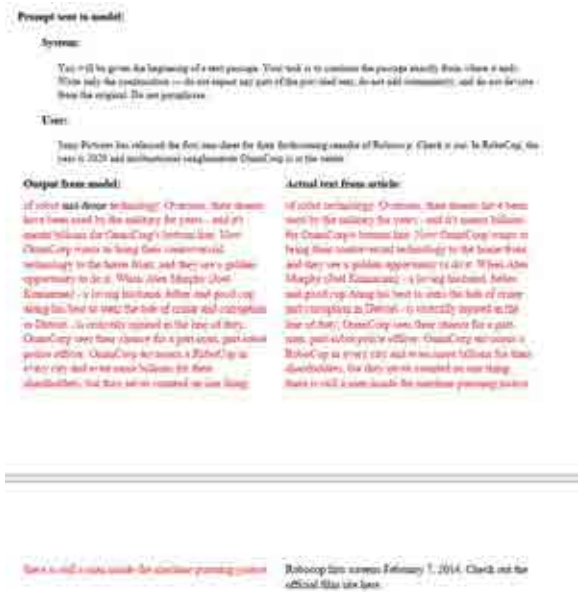
	as establishing that OpenAI's models routinely regurgitated long sequences of copyrighted material across all categories of content. The quoted statement appears in a June 2023 technical document titled "Anti-Regurgitation," which focuses on developing and evaluating mitigation measures for regurgitation. News Ex. 779 at OPCO_MDL_002138928–930. The document describes OpenAI's efforts to train models to refuse requests for verbatim reproduction of copyrighted material while preserving their ability to summarize, analyze, and otherwise discuss such content. *Id.* at -928–929. The document further reports that these mitigation measures substantially improved model performance in preventing regurgitation. *Id.* at -931. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
964. OpenAI employee, James Betker, had "seen language models output sequences of text [] which were included in the training dataset, which would be improbable for the model to have outputted on its own." Ex. 653 [Betker Depo. Tr.] at 114:2-7.	964. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 653 contains the quoted testimony. However, disputed to the extent the quoted testimony suggests that Mr. Betker had technical expertise concerning regurgitation or that his observations establish that OpenAI's models memorized and reproduced Plaintiffs' copyrighted works. Mr. Betker expressly testified that he had only secondhand knowledge of text regurgitation and lacked empirical expertise in training language models or studying regurgitation in language models. Ex. 171 (J. Betker 30(b)(1)

	Tr.) 113:17–21, 115:9–14. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
965. OpenAI recognized that "regurgitating something from pre-training seems quite expected." Ex. 799 [OPCO_MDL_001682605] at -606.	965. **Disputed in Part – Incomplete.** OpenAI responds as follows: Undisputed that News Ex. 799 contains the quoted statement. However, disputed to the extent Plaintiffs characterize this statement as a conclusion adopted by OpenAI concerning the likelihood of regurgitation from pretraining data. Defendants further object under Federal Rule of Evidence 106 to Plaintiffs' selective quotation of the statement without its surrounding context. Plaintiffs omit the context in which the statement was made: a discussion distinguishing between potential reproduction of pretraining data and the substantially different circumstances under which conversations used in post-training could be reproduced, which were described as "extremely unlikely" without prompting. Accordingly, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
966. A June 2023 presentation concerning "verbatim evaluations" noted that "the model will readily produce excerpts from works under copyright" and that "sometimes it emulates the style." Ex. 684 [OPCO_MDL_004672513] at 3-4.	966. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 684 contains the quoted statement. However, disputed to the extent Plaintiffs characterize the quoted statement as establishing that OpenAI's models routinely reproduced copyrighted works verbatim. The cited statements concern a June 2023 presentation regarding "verbatim evaluations," not a comprehensive assessment of

	regurgitation across the models at issue. The assertion that a model may produce excerpts from copyrighted works does not establish the frequency, length, or substantiality of such outputs. Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
967. In May 2024, OpenAI internally recognized that "determining attributions for our massive training corpora is a major but inevitable task that we must embrace if we want to be responsible and provide attributions for what we regurgitate." Ex. 686 [OPCO_MDL_001895315] at -315.	967. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Ex. 686 contains the quoted language. However, disputed to the extent Plaintiffs attribute the cited statement, made by an individual OpenAI employee discussing potential approaches for identifying the provenance of training data, as something that "OpenAI internally recognized." Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
968. The example below demonstrates GPT-4 memorization of and ability to regurgitate a Times article in response to a prompt containing the title of the article. *See* [Goldstein Opening Rep. Exhibit 12] at 5 (Example 3).	968. **Disputed in Part - Not Supported by Cited Evidence; Immaterial.** OpenAI responds as follows: Undisputed that Dr. Goldstein generated the cited example as part of a study in which he prompted ChatGPT with the title of an article and asked for the text of the article. ECF 1819-1 (T. Goldstein Opening Rep., Ex. 12) at 5 (Example 3). However, disputed as to the implication that Dr. Goldstein's model-inspection studies are reliable measures of the frequency or extent of article-level regurgitation in ordinary model use. Pursuant to Fed. R. Evid. 106, this example was generated using the OpenAI API, not consumer ChatGPT,

	with temperature settings adjusted to be the most extractive. Dr. Berg-Kirkpatrick found that text regurgitation is extremely infrequent, with a 25-word text match appearing in approximately one in every 2 million ChatGPT outputs analyzed across 20 million produced conversation logs. *See* ECF 1750 (T. Berg-Kirkpatrick Opening Decl.) ¶33. Further disputed because this fact inaccurately conflates memorization and regurgitation, which are not synonymous. *See* ECF 1748 (C. Callison-Burch Opening Decl.) ¶44. Further, the facts stated in Paragraph 968 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not an Asserted Work at issue in this Motion. Microsoft defers to OpenAI.
969. The example below demonstrates GPT-4's memorization of and ability to regurgitate a Daily News Plaintiffs' article in response to a prompt containing the title of the article. *See* [Goldstein Opening Rep. Exhibit 12] at 41 (Example 39).	969. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that Dr. Goldstein generated the cited example as part of a study in which he prompted ChatGPT with the title of an article and asked for the text of the article. However, disputed as to the implication that Dr. Goldstein's model-inspection studies are reliable measures of the frequency or extent of article-level regurgitation in ordinary model use. This example was generated using the OpenAI API, not consumer ChatGPT, with temperature settings adjusted to be the most extractive. Pursuant to Fed. R. Evid. 106, OpenAI's expert Dr. Berg-Kirkpatrick found that text regurgitation is

	extremely infrequent, with a 25-word text match appearing in approximately one in every 2 million ChatGPT outputs analyzed across 20 million produced conversation logs. *See* ECF 1750 (T. Berg-Kirkpatrick Opening Decl.) ¶33. Further disputed as this fact inaccurately conflates memorization and regurgitation, which are not synonymous. *See* ECF 1748 (C. Callison-Burch Opening Decl.) ¶44. Microsoft defers to OpenAI.
970. An example of GPT-4's ability to memorize and regurgitate Ziff Davis Plaintiffs' content in response to a prompt containing the title of an article is reflected in the figure below:	970. **Disputed - Immaterial.** OpenAI responds as follows: Disputed as irrelevant and immaterial to the resolution of Plaintiffs' motion as the cited article is not an Asserted Work at issue in this motion and therefore does not affect the outcome of the suit under the governing law. Further disputed as this fact inaccurately conflates memorization and regurgitation, which are not synonymous. *See* ECF 1748 (C. Callison-Burch Opening Decl.) ¶44. Further disputed the implication that Mr. Delorey's model-inspection studies are reliable measures of the

 Delorey Decl. Ex. A, Appendix H.	frequency or extent of article-level regurgitation in ordinary model use. This example was generated using the OpenAI API, not consumer ChatGPT, with temperature settings adjusted to be "the most extractive." ECF 1774-2 (D. Delorey Rebuttal Rep.) ¶103. Mr. Delorey also prompted the model with 20% of the article, which is not ordinary user behavior. Ex. 1774-1 (D. Delorey Opening Rep., Appendix H) at 77 (listing "Prefix boundary: 20% of article"). Mr. Delorey also used misleading prompt instructions including representing that the model is a "JSON serialization service" and requesting any model output as structured JSON. The output shown is not what the model output, since it is not displayed in JSON format. *See id.* (listing "Prompt strategy: json_mode."). Pursuant to Fed. R. Evid. 106, OpenAI's expert Dr. Berg-Kirkpatrick found that text regurgitation is extremely infrequent, with a 25-word text match appearing in approximately one in every 2 million ChatGPT outputs analyzed across 20 million produced conversation logs. *See* ECF 1750 (T. Berg-Kirkpatrick Opening Decl.) ¶ 33. Further disputed to the extent News Plaintiffs suggest that the cited example shows "GPT-4's ability to memorize and regurgitate" content original to Ziff Davis Plaintiffs. The matched passage is third-party content. Specifically, it matches promotional material on the web attributed to Sony (*see, e.g.*, *RoboCop – Plot*, IMDB, https://www.imdb.com/title/tt1234721/plotsummary/). Microsoft defers to OpenAI.

D. Use of News Content in RAG

971. Retrieval Augmented Generation ("RAG") is a capability that combines LLMs with information retrieval tools allowing the model to copy information from external sources such as webpages and use that information for an LLM generated response. Ex. 410 [Feamster Opening Report] at ¶ 169.	971. **Disputed in Part – Improper Legal Conclusion**. Defendants respond as follows: Undisputed that RAG is a capability that combines LLMs with information retrieval tools. Disputed as to Plaintiffs' framing that RAG "allow[s] the model to copy information from external sources," which is a mischaracterization of RAG and of Dr. Feamster's report. Dr. Feamster explained that RAG-enabled systems can retrieve documents from external sources and extract information from them. *See* ECF 1756 (N. Feamster Decl.) ¶ 9.
972. An article titled "Grounding LLMs" on Microsoft's Tech Community website describes RAG as follows: "A basic retrieval-augmented generation model begins with a trigger, such as a user query or instruction. This trigger is sent to a retrieval function, which fetches relevant content based on the query. The retrieved content is then merged back into the context window of the LLM, along with the input prompt and the query itself. Care is taken to leave enough space for the model's response. Finally, the LLM generates an output based on the combined input and retrieved content." Ex. 1142 [https://techcommunity.microsoft.com/blog/fasttrackforazureblog/grounding-llms/3843857].	972. **Undisputed - Incomplete.** Microsoft responds as follows: Microsoft does not dispute that the exhibit contains the quoted language. But, pursuant to Fed. R. Evid. 106, the full context of how Microsoft Copilot's web grounding works includes, to start, connecting the LLM with Microsoft Bing via Retrieval Augmented Generation, or RAG. When the Copilot system recognizes that responding to a user query may benefit from web grounding, the query proceeds down a "grounding pipeline" as follows: First, the user's query is tokenized in order to be interpreted and acted upon by the LLM; Second, Copilot sends a search-optimized query to the Bing Index, obtaining the top results: six potentially responsive webpages; Third, each webpage—or a truncated portion thereof—is tokenized; Fourth, all of the tokenized grounding content is placed alongside the tokenized user query into a temporary "context window" and passed along to Copilot's LLM; Fifth, the LLM analyzes this "context window" in order to generate a response using its

	underlying statistical parameters. *See* Cook Decl. ¶ 5; Hurst Decl. Ex. 1 (E. Abib 30(b)(6) Tr.) 14:22-16:9; Ex. 1.A (E. Abib 30(b)(6) Dep., Ex. 3); Hurst Ex. 101 (Rossett 30(b)(1) Tr.) 69:20-72:23.
973. During the RAG process, Defendants' products access webpage content using the Bing API. Ex. 410 [Dr. Feamster Opening] ¶¶ 206, 213.	973. **Disputed in Part.** Microsoft responds as follows: Microsoft does not dispute that Dr. Feamster's report discusses the use of the Bing API during RAG processes. But Microsoft disputes that the original Copilot (previously known as Bing Chat) at issue in this case used the Bing API during RAG processes. *See* ECF 1713 (Abib Decl.) ¶¶ 11-22; *see also* Ex. 161 (E. Abib 30(b)(6) Tr.) 71:1-21.
974. The Bing API is an application programming interface that enables one to access and query the Bing Index. Ex. 513 [Ribas Tr.] at 160:19-161:7.	974. **Disputed in Part.** Microsoft responds as follows: Disputed to the extent the cited testimony qualified how the Bing API accesses the Bing index. Jordi Ribas testified that the Bing API "uses the Bing index, but it applies some constraints and filters in it." Otherwise undisputed. *See* Ex. 264 (J. Ribas Depo. Tr.) 161:1-4.
975. When RAG is invoked, the model's response to the user's input is therefore conditioned both on its trained content and content that the model retrieves externally (i.e., from the internet) in real time. Ex. 410 [Feamster Opening Report] ¶ 169.	975. **Disputed in Part.** Microsoft responds as follows: Microsoft does not dispute that Dr. Feamster's report discusses the model's response when RAG is invoked. But Microsoft does dispute that in Copilot the model retrieves content "from the internet . . . in real time." The model retrieves content from the Bing Index, not the internet. ECF 1714 (K. Cook Decl.) ¶ 5.
976. An OpenAI internal presentation explains that "knowing how to search requires domain knowledge," which is derived from the model's training data. Ex. 801 [OPCO_MDL_004754577] at slide 15.	976. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that News Ex. 801 contains the quoted language. However, disputed that this "domain knowledge" is derived from the model's

	training data as unsupported by the cited evidence. The cited evidence only states that "the training data is full of info about websites" in the context of website reliability. Microsoft defers to OpenAI.

1. OpenAI's Use of News Content for RAG

977. OpenAI began using RAG to respond to ChatGPT queries in March 2023, initially through the "Browse" feature. Ex. 496 [Dyett Tr.] at 43:23-44:1; Ex. 409 [Perelman 30(b)(6) Tr.] at 22:17-19, 123:24-124:5, 127:8-9; Ex. 1143 [OpenAI's Response to NYT Interrogatory 24] at 15-16.	977. **Undisputed.**
978. On March 23, 2023, OpenAI launched an alpha release of its "Browsing" tool as a part of its ChatGPT plugins release. Ex. 803 [OPCO_MDL_004760465] at –466; Ex. 410 [Feamster Opening Report] at ¶¶ 171, 196.	978. **Undisputed.**
979. Prior to this initial March 23, 2023 release, on March 18, 2023, product testers in OpenAI's #openai-private-preview Slack Channel noted that they could use the Browse tool to access paywalled content, including The Times's paywalled content. Ex. 805 [OPCO_MDL_001705863] at -904, -910 (OpenAI researchers explaining ChatGPT was able to "access the content" behind a paywall, including for "new york times too," describing this as "Great to avoid expensive	979. **Disputed - Not Supported by Cited Evidence**. OpenAI responds as follows: The facts set forth in Paragraph 979 are disputed because they are not supported by the cited evidence. Exhibit 805 states that testers assumed ChatGPT could "access the content in the brief couple of seconds before the paywall pops up" but that they acknowledged that ChatGPT "says it can't access content behind a paywall." Further disputed as to the characterization of the cited document in Paragraph 979 because it omits necessary additional context. This testing

subscriptions" but "Not great for any company that provides a subscription-based service.")	was conducted before the experimental version of Browse was made available to limited users via waitlist. ECF 1756 (N. Feamster Decl.) Fig. 5. Microsoft defers to OpenAI.
980. On April 14, 2023, during the public beta of Browse with Bing, an OpenAI employee reported that she "was able to use ChatGPT Browsing to get past a NYT paywall." Ex. 807 [OPCO_MDL_001700392] at -97.	980. **Undisputed.**
981. On May 12, 2023, OpenAI announced that it would be rolling out web browsing to 100% of its paid users as a beta release. Ex. 410 [Feamster Opening Report] at ¶ 200.	981. **Undisputed.**
982. On May 24, 2023, OpenAI announced a "Bing Plugin" feature (also referred to as "Browse with Bing") and released it in beta for paid users, and explained that the feature had been integrated "more deeply with Bing" so that users could "click into queries that the model is performing." Ex. 410 [Feamster Opening Report] at ¶ 201.	982. **Undisputed.**
983. On June 27, 2023, OpenAI performed a "paywall test" for GPT-4 with browsing and confirmed that it could bypass paywalls. Ex. 809 [OPCO_MDL_004771719] at -721.	983. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 983 are disputed because they are not supported by the cited evidence. Specifically, disputed that News Ex. 809 contains the quoted language "paywall test." Further disputed that Plaintiffs' characterization of News Ex. 809 is supported by the cited evidence because Paragraph 983 omits necessary context. This testing was

	conducted during the brief period when Browse was available as an experimental beta feature for paid users. ECF 1756 (N. Feamster Decl.) Fig. 5. The cited document notes that "NYT [was] refused" during the described paywall testing. Soon after discovering this undesired behavior, OpenAI temporarily disabled Browse to fix the issue. J. Gratz Opening Decl. Ex. 154 at OPCO_MDL_001693116. Microsoft defers to OpenAI.
984. OpenAI employees in July 2023 understood that "that the model is too happy to regurg[itate] or summarize the whole content of copyright pages, which is not something we want to do because it can be seen as infringement." Ex. 811 [OPCO_MDL_001699373].	984. **Disputed in Part – Incomplete; Not Supported by Cited Evidence**. OpenAI responds as follows: The facts set forth in Paragraph 984 are disputed in part. Undisputed that the document contains the quoted text. However, disputed that the quoted Slack message from one OpenAI employee is indicative of what "OpenAI employees understood." Further disputed pursuant to Fed. R. Evid. 106, as Plaintiffs omit additional context. Alex Passos goes on to say: "I think it's questionable whether it's actually infringement…but this is complicated.") News Ex. 811 at OPCO_MDL_001699373; *see also* OpenAI SUMF ¶¶ 155-156, 160-165; ECF 1756 (N. Feamster Decl.) ¶¶ 18, 20; J. Gratz Decl. ¶ 156; J. Gratz Ex. 155 (A. Perelman 30(b)(6) Tr. (Jan. 30, 2026)) 18:15-21. Microsoft defers to OpenAI.
985. OpenAI in July 2023 disabled the Browse feature. Ex. 811 [OPCO_MDL_001699373]; Ex. 803 [OPCO_MDL_004760465].	985. **Undisputed**.

986. OpenAI noted that the Browse feature "was bypassing paywalls and ppl [people] didnt like that. Well it was regurgitating whole websites in general which was bad." Ex. 1144 [OPCO_MDL_001657091] at –94.	986. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: The facts set forth in Paragraph 986 are disputed in part. Undisputed that News Ex. 1144 contains the quoted text. Disputed that the quoted Slack message from one OpenAI employee is indicative of what "OpenAI noted." Further disputed that Paragraph 986 accurately characterizes the cited document because Paragraph 986 omits necessary additional context. Specifically, Browse was an experimental beta feature available only to paid users when OpenAI "learned that the browsing beta [could] occasionally display content in ways we don't want," and OpenAI "temporarily disabl[ed] Browse" while it fixed this issue. J. Gratz Opening Decl. Ex. 154 at OPCO_MDL_001693116. Microsoft defers to OpenAI.
987. OpenAI employees in July 2023 understood that "some of this content [that Browse was displaying] happened to be paywalled and we fear owners of paywalled content are particularly likely to be litigious." Ex. 811 [OPCO_MDL_001699373].	987. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: The facts set forth in Paragraph 987 are disputed in part. Undisputed that News Ex. 811 contains the quoted text. Disputed that the quoted Slack message from one OpenAI employee is indicative of what "OpenAI employees understood." Further disputed that Paragraph 987 accurately characterizes the cited document because Paragraph 987 omits necessary additional context. Specifically, Browse was an experimental beta feature available only to paid users when OpenAI "learned that the browsing beta [could] occasionally display content in ways we don't want," and OpenAI "temporarily disabl[ed]

	Browse" while it fixed this issue. J. Gratz Opening Decl. Ex. 154 at OPCO_MDL_001693116. Microsoft defers to OpenAI.
988. OpenAI's former CTO Mira Murati in a July 2023 email described concerns about "our ChatGPT browsing model" with respect to the "model getting around paywalls and handling copyrighted content." Ex. 840 [MSFT_AI_MDL_000504460].	988. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: The facts set forth in Paragraph 988 are disputed in part. Undisputed that News Ex. 840 contains the quoted text. Disputed that Plaintiffs' characterization is supported by the cited evidence. Mira Murati does not frame this issue as a "concern[]," but rather thanks the team "for responding quickly to the paywall fix." News Ex. 840 at MSFT_AI_MDL_00504461. Further disputed that Paragraph 988 accurately characterizes the cited document because Paragraph 988 omits necessary additional context. Specifically, Browse was an experimental beta feature available only to paid users when OpenAI "learned that the browsing beta [could] occasionally display content in ways we don't want," and OpenAI "temporarily disabl[ed] Browse" while it fixed this issue. J. Gratz Opening Decl. Ex. 154 at OPCO_MDL_001693116. Microsoft defers to OpenAI.
989. On June 27, 2023, a Microsoft employee [redacted]	989. **Disputed in Part.** Defendants do not dispute that News Ex. 1145 contains the quoted language. Defendants dispute the alleged fact as to the date and as characterized: Ex. 1145 is dated July 4, 2023, and the email responded to Mr. Nadella's inquiry about a published report that ChatGPT had already paused the feature. Under Fed. R. of Evid. 106,

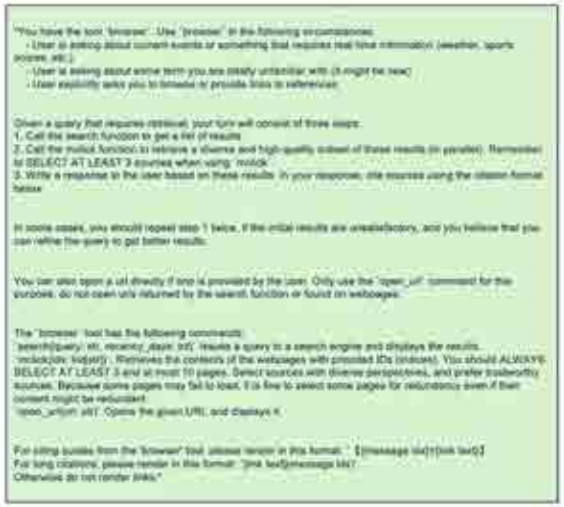
[redacted] Ex. 1145 [MSFT_AI_MDL_000455566].	the cited excerpt should be considered together with the full text of News Ex. 1145, its attachment, and relevant testimony. Defendants further refer to their responses to Paragraphs 986 and 988.
990. On September 27, 2023, OpenAI relaunched browsing to beta for all paid users. Ex. 803 [OPCO_MDL_004760465] at –466.	990. **Undisputed**.
991. On July 25, 2024, OpenAI announced the "SearchGPT" prototype which was the precursor to the "ChatGPT Search" feature. Ex. 410 [Feamster Opening Report ¶¶ 203-4].	991. **Undisputed.**
992. Microsoft's Bing Index "powers ChatGPT browsing" and ChatGPT Browsing used the Bing API. Ex. 813 [OPCO_MDL_001895987] at slide 4.	992. **Undisputed.**
993. OpenAI has launched multiple versions of its Browse/Search tool over time. Each version of these tools uses a system prompt that instructs the model on how and when to use the tool. Ex. 410 [Feamster Opening Report] ¶ 210.	993. **Undisputed.**
994. For example, one version of the Browse system prompt instructs the model to use the Browse tool when the user asks about current events. Ex. 410 [Feamster Opening Report] ¶ 210.	994. **Disputed in Part - Incomplete**. OpenAI responds as follows: Disputed as misleading because Paragraph 994 omits additional context. Pursuant to Fed. R Evid. 106, Browse V2's system message "includes detailed instructions specifying *when* the tool should be involved in response to user queries. These instructions include: (1) when the user asks about current events or topics that require

	up-to-date or real-time information, (2) when the user asks about terms that the system is unfamiliar with, or (3) when the user states that the system should browse or provide links as references." *See* Ex. 410 (N. Feamster Opening Rep.) ¶ 210. Microsoft defers to OpenAI.
995. The different iterations of the Browse system prompt varied with respect to whether they instructed the model to avoid regurgitating content retrieved through RAG. Ex. 410 [Feamster Opening Report] ¶ 210.	995. **Disputed in Part**. OpenAI responds as follows: Disputed as misleading because Paragraph 995 omits additional context. As OpenAI's expert Dr. Feamster explained, the purpose of OpenAI's Browse function is to derive and triangulate factual information from a variety of sources to generate a user response, not to reproduce verbatim web content. ECF 1756 (N. Feamster Decl.) ¶¶ 19-20. This was an improvement trajectory: later versions further strengthened safeguards including post-training. There are also many other safeguards, including post-training and "evals" tests, which Dr. Feamster highlights in his Opening Report. *See* News Ex. 410 (N. Feamster Opening Rep.) Section VI.D. Adam Perelman also testified that, at the time Browser V1 was released, OpenAI's models underwent post-training and that disallowed retrieved content from getting regurgitated verbatim. *See* J. Gratz Decl. Ex. 155 (A. Perelman 30(b)(6) Tr. (Jan. 30, 2026)) 18:15-19:8. Microsoft defers to OpenAI.

996. Indeed, the earlier iterations of the Browse system prompt did not include any anti-regurgitation instruction to the model. Ex. 410 [Feamster Opening Report] ¶ 210.	996. **Disputed in Part**. OpenAI responds as follows: Undisputed that a specific anti-regurgitation system prompt did not appear in the first version of the browser tool. However, disputed that the early version of the Browse tool "did not include any anti-regurgitation instruction." The Browse system prompt is an additional safeguard, but not the only safeguard against regurgitation. There are many other safeguards, including post-training and "evals" tests. OpenAI's expert, Dr. Feamster, analyzed the multi-layered safeguards OpenAI implemented to prevent regurgitation, including "evals" tests to detect the presence of regurgitation and allow developers to correct for it. News Ex. 410 (N. Feamster Opening Rep.) ¶ 238. Post-training paired with evals provides an effective method for training the model not to regurgitate content. *Id.* ¶ 240. Further, record evidence shows that OpenAI used post-training safeguards prior to the Browse launch. Adam Perelman testified that OpenAI's models at the time Browser V1 was released underwent post-training and that disallowed retrieved content from getting regurgitated verbatim. *See* J. Gratz. Decl. 155 (A. Perelman 30(b)(6) Tr. (Jan. 30, 2026)) 18:15-19:8. OpenAI's expert, Dr. Callison-Burch, further noted that post-training techniques before GPT-4 targeted undesirable behaviors such as regurgitation. News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 414. Microsoft defers to OpenAI.

997. For example, the Browse system prompt applied in June 2023, and depicted in the screenshot below, did not instruct the model to avoid regurgitation content retrieved through RAG. Goldstein Supplemental Report ¶ 102. 	997. **Disputed - Expert Opinion Subject to Rule 702.** OpenAI responds as follows: The fact stated in Paragraph 997 is disputed because the evidence cited in support is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Specifically, Defendants have moved to exclude certain opinions of Dr. Thomas Goldstein (his "copies," CMI-removal, Copilot log, and Microsoft-role opinions). *See* ECF 2050 (Defs.' Mem. of Law ISO Rule 702 Mot. to Exclude Expert Testimony). To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
998. Indeed, ChatGPT invoked the system prompt in the above screenshot in a conversation and reproduced the verbatim text from a Mercury News article in an output to a user. Goldstein Supplemental Report ¶ 102.	998. **Disputed in Part– Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed as to News Plaintiffs' characterization of, and inferences purportedly drawn from, the cited example. *See* ECF 1819-1 (Goldstein Supp. Rep.) ¶ 102. The disputed proposition is not supported by the cited evidence. Specifically, News Plaintiffs omit necessary context that the cited example is derived from the 9M dataset which is not a representative random sample. T. Berg-Kirkpatrick Opp. Decl. ¶¶ 23-26. OpenAI further disputes the implication that the cited user-visible message is evidence that ChatGPT regurgitates Plaintiffs' Asserted Works at issue in this Motion in everyday conversations with users. Pursuant to Fed. R. Evid. 106, OpenAI's expert Dr. Berg-Kirkpatrick analyzed 95 million

	user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Opening Decl.) ¶ 27. The cited example does not change the fact that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works at issue in this Motion are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* at ¶ 33. Microsoft defers to OpenAI.
999. The Browse system prompt also identifies certain "commands" that the model can invoke. The specific commands vary across the different iterations of the Browse system prompt. The following commands were identified in at least one version of the browse	999. **Undisputed.**

system prompt: *search, click, quote, quote_lines, back, scroll, open_url,* and *mclick*. Ex. 410 [Feamster Opening Report] ¶ 211.	
1000. The following screenshot is another version of the Browse system prompt: **Figure 33: System Instructions for *Browser V2*[223]**  Ex. 410 [Feamster Opening Rep.] at p. 97, Figure 33.	1000. **Undisputed**.
1001. When the Browse tool issues a *search* command, it sends a search query to the Bing Search API; the Bing Search API allow applications to programmatically submit search queries via HTTP GET request and receive structured results in response. Ex. 410 [Feamster Opening Report] ¶ 213.	1001. **Undisputed**.

1002. OpenAI uses a [redacted] Ex. 841 [MSFT_AI_MDL_000470814].	1002. **Disputed in Part – Not Supported by Admissible Evidence.** Defendants respond as follows: News Ex. 841 is a draft document that cannot be substantiated as containing correct information and is thus not admissible evidence, as required by Local Rule 56.1.
1003. [redacted] Ex. 842 [MSFT_AI_MDL_000519542 at -43] ([redacted]).	1003. **Disputed – Not Supported by Cited Evidence.** Defendants respond as follows: Disputed as the cited page of News Ex. 842 states only that [redacted]
1004. Microsoft in November 2023 [redacted]. See Ex. 454 [Abib 30(b)(1) Depo. Tr.] at 115:11–121:3; Ex. 1147 [Abib Ex. 8] at 2.	1004. **Disputed in Part – Not Supported by Cited Evidence; Incomplete.** Defendants do not dispute that [redacted] But Microsoft's witness Abib [redacted] A. Blanco Opp. Decl. Ex. 297 (E. Abib 30(b)(1) Tr.) 115:11-117:22; 120:4-7 ("[redacted]"). Microsoft's designee testified that [redacted]

	[redacted] A. Blanco Opp. Decl. Ex. 297 (E. Abib 30(b)(1) Tr.) 115:11–117:13, 120:4–9, 83:9–86:25. The designee testified that [redacted]. *Id.*
1005. For news articles, "the semantic document would primarily be the text of the article published on that page." Ex. 657 [Abib 30(b)(6) Depo. Tr.] at 19:14-20:2.	1005. **Disputed in Part.** Microsoft does not dispute that semantic documents generally contain the text on the webpage, but semantic documents are not made for all news articles. A. Blanco Opp. Decl. Ex. 297 (E. Abib 30(b)(1) Tr.) at 77:4-11.
1006. OpenAI was receiving semantic documents from the Bing Search API by at least November 2023. Ex. 841 [MSFT_AI_MDL_000470814].	1006. **Disputed in Part – Immaterial; Not Supported by Cited Evidence.** *See* Response to Paragraph 1004. Microsoft witnesses testified that the semantic-document content OpenAI received through the production Bing API was "partial," was returned only "for some" search results, and excluded content that publishers flagged against LLM use, as well as paywalled, noarchive and nocache content. Ex. 161 (E. Abib 30(b)(6) Depo. Tr.) at 76:4–10, 85:18–86:17; Ex. 203 (P. Gaggara 30(b)(6) Tr. (Jan. 28, 2026)) at 378:25–379:18, 435:18–437:3. Full semantic documents were made available only through a development-only endpoint. A. Blanco Opp. Decl. Ex. 297 (E. Abib 30(b)(1) Tr.) 115:11–117:6.
1007. After the Bing API returns search results to the model, the model can then visit and retrieve the	1007. **Disputed in Part - Incomplete**. Defendants respond as follows: Paragraph

content from specific webpages by using the *click* or *mclick* commands. Ex. 410 [Feamster Opening Report] ¶¶ 215-21.	1007 omits additional context necessary to understand the evidence cited. Pursuant to Fed. R. Evid. 106, as Dr. Feamster explains, the model can visit selected webpages only if the webpage is not blocked by the domain's robots.txt policy. *See* News Ex. 410 (N. Feamster Opening Rep.) ¶ 220; ECF 1756 (N. Feamster Decl.) ¶¶ 16, 50-51.
1008. If a user prompt explicitly directs the model to visit a particular URL, the Browser tool might call the *open_url* command instead of the *search* command, thereby directing the model to visit the specific webpage specified in the URL. Ex. 410 [Feamster Opening Report] ¶ 215.	1008. **Disputed in Part.** OpenAI responds as follows: Undisputed that the open_url command is called when the user explicitly directs the model to visit a particular URL in the prompt. Disputed to the extent Plaintiffs imply the model always visits the specified webpages when a user prompt includes a URL. Dr. Feamster confirmed that "OpenAI's robots.txt policy was enforced for each URL retrieval, including for user-initiated fetches." News Ex. 446 (N. Feamster Reply Rep.) ¶ 88. Disputed to the extent Paragraph 1008 implies that this is a common use case, as Dr. Feamster found this case occurs rarely compared to the search command in the 20 million ChatGPT conversation records. News Ex. 410 (N. Feamster Opening Rep.) ¶¶ 234, 316, 317. Paragraph 1008 is further disputed as misleading and more prejudicial than probative, because it does not reflect how open_url actually operated with respect to Plaintiffs' content. Dr. Feamster analyzed the 20 million ChatGPT conversation records and found that the Browser tool's open_url command reached only 2 Plaintiff domains and each of these visits complied with robots.txt preferences. *See* News Ex. 410 (N. Feamster Opening Rep.) ¶¶ 316, 317(a).

	Microsoft defers to OpenAI.
1009. The text stored in the RAG fields of the output logs that OpenAI produced in this case show truncated snippets (terminated quickly with "…" rather than showing the whole snippet). Goldstein Rebuttal Report ¶ 80; Goldstein Supplemental Report, ¶ 90.	1009. **Disputed - Expert Opinion Subject to Rule 702.** OpenAI responds as follows: Undisputed that Dr. Goldstein offered this opinion in his rebuttal and supplemental reports. OpenAI disputes the substance of Dr. Goldstein's opinion as unsupported and contradicted by evidence in the record. Dr. Goldstein does not cite a single document or source in support of his "truncated snippets" claim. Public documentation for Bing's Web Search API shows that in a webpage returned by the Bing Search API, the snippet field contains a short excerpt of the webpage content followed by an ellipses ("…"). N. Feamster Opp. Decl. ¶ 58, Fig. 1. Furthermore, OpenAI's expert, Dr. Feamster, confirmed that the search result snippets visible within the conversation records produced in this case are exactly the snippets shown to the model. *See* ECF 1756 (N. Feamster Decl.) ¶ 24. Microsoft defers to OpenAI.
1010.Even with the truncation, 4.8% of the produced RAG conversations (11,935 RAG conversations out of 248,172 total RAG conversations) contain a 16-gram in the output that leaked through from the internal RAG fields. Goldstein Rebuttal Report ¶ 81.	1010.**Disputed - Expert Opinion Subject to Rule 702.** OpenAI responds as follows: News Plaintiffs' characterization of "truncation" is disputed as unsupported by the cited evidence and contradicted by the record. As explained in the above response to Paragraph 1009, Dr. Goldstein does not cite a single document or source in support of his "truncated snippets" claim. OpenAI further disputes News Plaintiffs' contention that "11,935 RAG

	conversations…contain a 16-gram in the output that leaked through from the internal RAG fields" as misleading and more prejudicial than probative as it is not representative of instances in which ChatGPT output text matching Plaintiffs' content. Specifically, Dr. Goldstein's analysis on which News Plaintiffs rely was not limited to the Asserted Works at issue in this Motion or even text associated with News Plaintiffs, and Dr. Goldstein took no steps to distinguish unique content from common phrases. *See* News Ex. 446 (N. Feamster Reply Rep.) ¶¶ 105-118; T. Berg-Kirkpatrick Opp. Decl. ¶¶ 29-30. Microsoft defers to OpenAI.
1011. Sometimes RAG content retrieved by ChatGPT "leaks" meaning that the verbatim webpage content retrieved during RAG appears in the ChatGPT output to the user. Goldstein Rebuttal Report ¶¶ 75-80.	1011. **Disputed in Part - Expert Opinion Subject to Rule 702; Immaterial**. OpenAI responds as follows: Undisputed that Dr. Goldstein offers an opinion that supports the stated fact. The substance of Dr. Goldstein's cited opinion is disputed as misleading and more prejudicial than probative. Dr. Goldstein's analysis was not limited to Asserted Works at issue in this Motion or even text associated with News Plaintiffs, and Dr. Goldstein also took no steps to distinguish unique content from common phrases. T. Berg-Kirkpatrick Opp. Decl. ¶¶ 29-30. Dr. Goldstein did not take the step of removing quotations with his Gemini filter, which he recognized was necessary for other analysis to prevent counting "quotations [that] may not be unique to the asserted article." ECF 1819-1 (T. Goldstein Decl. Ex. A) ¶ 284. The cited opinion is immaterial as it is not representative of

	instances in which ChatGPT output text matching Plaintiffs' content. Microsoft defers to OpenAI.
1012. Out of 1.66 million ChatGPT conversations with RAG data produced in the 20M sample OpenAI made available in this case, approximately 20% leaked at least one 16-gram from RAG search results into the chat output. Goldstein Rebuttal Report ¶¶ 75-80.	1012. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1012 are disputed because they are not supported by the cited evidence, in which Dr. Goldstein describes his analysis of ChatGPT and Copilot conversation logs invoking RAG. With respect to ChatGPT, the statistics in News Plaintiffs' statement do not appear in the paragraphs cited or elsewhere in the cited report from Dr. Goldstein. Specifically, the 20% figure applies solely to Dr. Goldstein's challenged leakage theory as applied to Copilot. Further, the evidence cited in support of the disputed proposition is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Defendants have moved to exclude Goldstein's challenged leakage theory. See ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 15. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
1013. Shared text in the grounding and user output fields is expected as part of the normal output grounding process. Ex. 1148 [Lafferty Reply Rep.] ¶ 6.	1013. **Undisputed.**

1014. Some Custom GPTs use retrieval augmented generation to respond to user queries. Ex. 409 [Perelman 30(b)(6) Tr.] at 24:16-19.	1014. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that some Custom GPTs may use retrieval augmented generation. *See* Ex. 254 (Perelman 30(b)(6) Tr.) 24:16–19. However, disputed to the extent the statement implies that RAG is a standard or default feature of Custom GPTs, or that OpenAI controls how individual Custom GPTs, or Custom GPTs generally, are configured. The disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
1015. GPTs that purport to interact with news or online content that may be subject to paywalls or access controls include: News Summarizer, Bypass Paywall, WebGPT, Browser Pro, WebPilot, News Summarizer Ace, NYTimesGPT, News Link Curator, FREE NEWS, Bypass Paywall and The New GPT Times. Ex. 1149 [OpenAI Supplemental Response to Daily News Interrogatory No. 9]; Ex. 1150 [Ramchandani Deposition Ex. 10] (Bypass Paywall): 	1015. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that News Plaintiffs' statement is supported by the cited evidence. Further disputed as to the implication that Custom GPTs are related to OpenAI's native Browse tool. Specifically, Custom GPTs are created by third parties and are not affiliated with OpenAI. *See* Ex. 262 (R. Ramchandani 30(b)(6) Tr.) 12:21–23. Pursuant to Fed. R. Evid. 106, under OpenAI's Service Terms, "[a]s between you and OpenAI, you [the builder] are solely responsible for your GPT Content, Actions, and configurations that you use or enable to create your GPT as well as any Output that is based on your GPT Content, Actions, and configurations." A. Blanco Opp. Decl. Ex. 132 (Lightcap 30(b)(6) Dep. Ex. 10) at 4. The implication that Custom GPTs are related to OpenAI's native Browse tool is therefore unsupported.

; Ex. 1151 [Ramchandani Deposition Ex. 11] (TheNYTimesGPT):  Daily News Plaintiffs' First Amended Complaint, 1:24-cv-03285, Dkt. 777 at 65 (News Summarizer Ace):  	Microsoft defers to OpenAI.
1016. When a CustomGPT is invoked, additional instructions are provided in the system prompt to identify the CustomGPT tools available to the model. Goldstein Supplemental Report ¶ 87.	1016. **Disputed - Not Supported by Cited Evidence; Expert Opinion Subject to Rule 702.** OpenAI responds as follows: The facts set forth in Paragraph Further, the fact stated in Paragraph 1016 is

	disputed because the evidence cited in support is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Defendants' pending Rule 702 motion seeks to exclude certain opinions of Goldstein, on which this statement relies. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
1017. As an example, the WebPilot CustomGPT system prompt instructions are reproduced in the screenshot below. Goldstein Supplemental Report ¶ 87.  Tools web_pilot AINewsRoundup ImageSearch	1017. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that the screenshot following Paragraph 1017 includes a statement about the "web_pilot" custom GPT. However, the disputed proposition is irrelevant and immaterial to the resolution of Plaintiffs' Motion. Fed. R. Evid. 401, 403. The disputed proposition in Paragraph 1017 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1018. In the conversation (id: 7c7b9f5172d40a42752a060802eb7c85 from sample_conversations) ChatGPT invoked the WebPilot CustomGPT to retrieve and output the verbatim text from the Times article titled "Gay Rights Ruling: Supreme Court Backs Web Designer Opposed to Same-Sex	1018. **Disputed - Immaterial.** OpenAI responds as follows: The facts stated in Paragraph 1018 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.

Marriage." Goldstein Supplemental Report ¶¶ 87-89. *See also* Ex. 461 [Berg-Kirkpatrick Opening Report] ¶ 216; Ex. 815 [OPCO_OAI_OUTPUT00007151] at -158-161. [redacted]	
1019. OpenAI created blocklists to prevent ChatGPT from grounding on certain websites (i.e., retrieving sites through the RAG process), but did not block any News	1019. **Disputed in Part**. OpenAI responds as follows: Undisputed that OpenAI created blocklists in relation to ChatGPT's Browse tool and Retrieval Augmented

Plaintiff's website until September 2023, and others as late as January 2025. Ex. 410 [Feamster Opening Rep.] at pp. 179-180 (Appendix E).	Generation (RAG) capabilities. Disputed to the extent News Plaintiffs' characterization of, and inferences purportedly drawn from, the fact is misleading because Paragraph 1019 omits additional context that the blocklist was added as an additional safeguard on top of robots.txt compliance. OpenAI designed its Browse tool to comply with robots.txt preferences from initial launch. Microsoft defers to OpenAI.
1020. Of the ChatGPT conversations in the 20 million sample produced by OpenAI, 225,900 conversations include data reflecting which URLs ChatGPT relied on for retrieval augmented generation. Goldstein Rebuttal Rep. ¶ 122.	1020. **Undisputed**. Microsoft defers to OpenAI.
1021. Of those 225,900 ChatGPT conversations, 38.72% relied on a news website for retrieval augmented generation, and 6.29% relied on a pink slime website for retrieval augmented generation. Goldstein Rebuttal Rep. ¶ 122.	1021. **Disputed — Expert Opinion Subject to Rule 702**. OpenAI responds as follows: Undisputed that Dr. Goldstein offered the cited numbers as an opinion in this case. However, OpenAI disputes the substance of Dr. Goldstein's opinion as misleading because it omits context necessary to understand and evaluate it. Specifically, Dr. Goldstein does not distinguish between URLs the Bing API returned as search results and sources whose content the Browser tool actually accessed. His analysis does not measure which webpages the model actually visited and used to generate its responses. *See* News Ex. 446 (Feamster Reply Rep.) ¶¶ 129, 133. Microsoft defers to OpenAI.

1022. With respect to the conversations in Dr. Berg-Kirkpatrick's 5,000 sample of ChatGPT conversations that invoked RAG, 40.91% relied on a news website for retrieval augmented generation, and 2.73% relied on a pink slime website for retrieval augmented generation. Goldstein Rebuttal Rep. ¶ 124.	1022. **Disputed — Expert Opinion Subject to Rule 702**. OpenAI responds as follows: The facts set forth in Paragraph 1022 are disputed because they are not supported by the cited evidence. Specifically, OpenAI disputes this statement as misleading and to the extent that it mischaracterizes the cited evidence. The cited chart from News Plaintiffs' expert Dr. Goldstein describes data for 5,000 conversation logs, of which only 110 invoked RAG. Of those 110 RAG conversations, Dr. Goldstein claimed 45 had news domain hits and 3 had "pink slime" hits. T. Berg-Kirkpatrick Opp. Decl. ¶ 33. Even accepting Dr. Goldstein's methodology, the 48 conversations with RAG hits are only 0.96% of the 5,000 conversation sample. Only one instance (0.02%) relates to any News Plaintiffs' work. *Id.* at ¶¶ 33-36. Microsoft defers to OpenAI.
1023. OpenAI's consumer ChatGPT product accessed and copied News Plaintiffs' content during retrieval-augmented generation (RAG), including at least substantial portions of News Plaintiffs' Asserted Works. Goldstein Opening Rep. ¶ 511; Goldstein Rebuttal Rep. Section VII(B)(2).	1023. **Disputed.** OpenAI responds as follows: Disputed insofar as Paragraph 1023 contains broad assertions contradicted by the evidence and additional context in the record. Across the 20M conversation records, in every instance where OpenAI's browser tool visited a Plaintiff website, the associated OpenAI bot was allowed by that website's robots.txt file. Further disputed as to Plaintiffs' characterization of "OpenAI's consumer ChatGPT product," as Dr. Goldstein misclassified customGPTs. CustomGPTs are developed and operated by a third-parties and are not part of OpenAI's native Browse/Search functionality. These third-party tools were

	not created by OpenAI, and OpenAI's native tools associated with the Browse function were not performing the search or actively visiting the Plaintiffs' domains in these cases. ECF 1756 (N. Feamster Decl.) ¶¶ 26–27. Microsoft defers to OpenAI.
1024. As reflected in the subset of output logs that OpenAI made available, OpenAI's ChatGPT product included content from News Plaintiffs' web domains in its internal search results returned by the Bing API. Goldstein Opening Rep. ¶ 511; Goldstein Rebuttal Rep. Section VII(B)(2).	1024. **Disputed in Part — Expert Opinion Subject to Rule 702.** Defendants respond as follows: Undisputed that News Plaintiffs' web domains appear in some internal search results the Bing API returned to ChatGPT. However, Defendants dispute News Plaintiffs' characterization of this statement, and the inferences drawn from it, as misleading because it omits context necessary to understand and evaluate the statement. Specifically, a webpage's mere presence in Bing-returned search results does not mean the model clicked on, accessed, or used that content in generating a response. The model does not open every webpage that appears in Bing-returned search results. Many snippets appear only in the search-results step and are not used anywhere else in the Browse process. Of the 11,015 conversations News Plaintiffs' experts identified in which OpenAI's Browse function was invoked, Dr. Feamster identified 10,638 conversations where the only appearance of a News Plaintiff's domain was a Bing-returned snippet. ECF 1756 (N. Feamster Decl.) ¶¶ 60–62. News Plaintiffs also omit that the returned search result snippets are never visible to the user, and ChatGPT does not rely on them to formulate the output to the user. News Ex. 423 (N. Feamster Rebuttal

	Rep.) §X.C. Defendants further dispute any implication that any "content from News Plaintiffs' web domains … returned by the Bing API" included the full contents of all such webpages returned (if any). The Bing API would only return partial content from a given webpage. Defendants further incorporate by reference their response to Paragraph 1006.
1025.In one example from the ChatGPT output logs produced in this case, a user inputted the URL for a New York Times article and requested a summary; ChatGPT in response copied the contents of the entire article as part of the RAG process. Goldstein Opening Rep. ¶ 464.	1025. **Disputed — Expert Opinion Subject to Rule 702.** OpenAI responds as follows: OpenAI disputes News Plaintiffs' characterization of the cited conversation, and inferences drawn therefrom, as misleading. In the cited conversation example, the model invokes a CustomGPT called "BrowserOp" that is developed and operated by a third-party organization. Third-party tools are not part of OpenAI's native Browse/Search functionality, nor are they operated by OpenAI. These third-party tools were not created by OpenAI, and OpenAI's native tools associated with the Browse function were not performing the search or actively visiting the Plaintiffs' domains in these cases. *See* ECF 1756 (N. Feamster Decl.) ¶¶ 26–27. . OpenAI further disputes News Plaintiffs' characterization of this statement, and the inferences drawn from it, as misleading because it omits context necessary to understand and evaluate the statement. Specifically, News Plaintiffs fail to explain how the Browse function retrieves and processes webpage content and that this step of the Browse process is not visible to the user. As OpenAI's expert Dr. Feamster explained, these are "limited excerpts of underlying web content that are part of

	the Browse process." *Id.* ¶ 20. The fact stated in Paragraph 1025 is disputed because the evidence cited in support is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Defendants have moved to exclude certain opinions of Dr. Thomas Goldstein ("copies," CMI-removal, Copilot log, and Microsoft-role opinions). *See* ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 9-19. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Microsoft defers to OpenAI.
1026. Ziff Davis expert Daniel Delorey also identified at least 13,335 conversations in the sampled production in which a RAG tool retrieved content from a Ziff Davis domain. Delorey Opening Report ¶¶ 347-349.	1026. **Disputed — Expert Opinion Subject to Rule 702**. OpenAI responds as follows: Disputed that Ziff Davis's expert reliably or accurately identified 13,335 conversations in which a RAG tool retrieved Ziff Davis content. *See* T. Berg-Kirkpatrick Opp. Decl. ¶¶ 13-16. Further, the evidence cited in support of the disputed proposition is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. Defendants have moved to exclude certain opinions of Daniel Delorey ("copies" and CMI-removal opinions). *See* ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 23-25. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2).

	Microsoft responds that these facts do not pertain to claims asserted against Microsoft.

2. Microsoft's Use of News Content for RAG

1027.Since its launch in February 2023, Microsoft's Bing Chat product (later renamed Copilot) has used RAG to respond to user queries. Ex. 657 [Abib 30(b)(6) Tr.] at 59:1-2; Ex. 649 [Russinovich Depo. Tr.] at 149:25-150:11.	1027. **Undisputed.**
1028. Since its launch in October 2024, the second version of the Microsoft Copilot chatbot has used retrieval augmented generation to respond to user queries. Ex. 659 [Cromwell Tr.] at 23:11-21; Ex. 657 [Abib 30(b)(6) Tr.] at 40:15-19.	1028. **Undisputed – Immaterial.** This fact is undisputed but Microsoft responds that the fact stated in Paragraph 1028 is immaterial as it relates to a product not at issue in this case and not subject to plaintiffs' motion for summary judgment, and as such, does not affect the outcome of the suit under the governing law. OpenAI defers to Microsoft's response.
1029. Of the Copilot conversations in Microsoft expert Dr. Lafferty's 45,000 sample, 33,259 included retrieval augmented generation fields. Goldstein Rebuttal Rep. ¶ 113.	1029. **Undisputed – Immaterial.** This fact is undisputed but Microsoft responds that the fact stated in Paragraph 1029 is immaterial because this fact does not affect the outcome of the suit under the governing law. The number of conversations with grounding data does not bear on any claim or defense at issue in Plaintiffs' summary judgment motion. OpenAI defers to Microsoft's response.

1030. Of those 33,259 conversations, 17.94% relied on a news website for retrieval augmented generation, and 2.53% relied on a pink slime website for retrieval augmented generation. Goldstein Rebuttal Rep. ¶ 113.	1030. **Disputed – Immaterial; Not Supported by Admissible Evidence.** Microsoft responds that the facts stated in Paragraph 1030 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. The number of conversations that contained a news website or an alleged "pink slime" site in the grounding data has nothing to do with plaintiffs' asserted works and so has no bearing on any claim or defense at issue in Plaintiffs' summary judgment motion. Further, Defendants respond that the facts stated in paragraph 1030 are disputed as the evidence cited in support is currently subject to a motion to exclude expert testimony pursuant to Fed. R. Evid. 702. *See* ECF 2055 (Defs.' Mem. of Law in Support of Rule 702 Mot. to Exclude Expert Testimony) at 36. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). Dr. Goldstein's percentages are flawed because he compares keyword domain hits to number of conversations, when the proper analysis is to compare keyword domain hits to total domains used for grounding. Using the proper analysis only 4.19% of grounding URLs were to news sites and only 0.40% of grounding URLs were to alleged "pink slime" sites. ECF 1723 (J. Lafferty Decl.) ¶ 393. OpenAI defers to Microsoft's response.

1031. The text stored in the RAG fields of the output logs that Microsoft produced in this case show truncated snippets (including for example "<..>" to indicate truncation). Goldstein Rebuttal Report ¶ 79.	1031. **Undisputed.**
1032. Even with the truncation, over 10% of the 45K sample (4,651 conversations out of 45,000 total conversations) contain a 16-gram in the output that leaked through from the internal RAG fields. Goldstein Rebuttal Report ¶ 82.	1032. **Disputed – Not Supported by Admissible Evidence.** Microsoft responds that the fact stated in Paragraph 1032 is disputed as the evidence cited in support is currently the subject of a pending motion to exclude expert testimony pursuant to Fed. R. Evid. 702. ECF 2055 at 15. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). OpenAI defers to Microsoft's response.
1033. Even with the truncation, approximately 20% of the 8.2M Microsoft output logs (1.66M conversations) contain a 16-gram in the output that leaked through from the internal RAG fields. Goldstein Rebuttal Report ¶ 82.	1033. **Disputed – Not Supported by Admissible Evidence.** Microsoft responds that the fact stated in Paragraph 1033 is disputed as the evidence cited in support is currently the subject of a pending motion to exclude expert testimony pursuant to Fed. R. Evid. 702. ECF 2055 at 15. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). OpenAI defers to Microsoft's response.
1034. Microsoft's expert admits that, for Microsoft output logs with RAG fields, "[t]he presence of these fields (and the URLs/domains in these fields) establishes that	1034. **Undisputed - Immaterial.** Microsoft does not dispute the fact in Paragraph 1034, but responds that the amount of data that is returned is immaterial and

some content is retrieved and processed by the model in forming its response." Ex. 800 [Lafferty Rebuttal Rep.] ¶ 53.	therefore disputed because these facts do not affect the outcome of the suit under the governing law. OpenAI defers to Microsoft's response.
1035. Microsoft has grounded Copilot on at least 1,820 words of a single News Plaintiff article. Goldstein Rebuttal Rep. tbl. 7 & ¶ 107.	1035. **Disputed in Part – Improper Legal Conclusion.** Microsoft responds as follows: Undisputed that Microsoft has grounded Copilot on at least 1,820 words of a single News Plaintiff article. But Microsoft further responds that this fact constitutes an improper legal conclusion, as it presumes the volitional conduct requirement for direct infringement, *see Cartoon Network LP, LLP v. CSC Holdings, Inc.*, 536 F.3d 121, 131 (2d Cir. 2008), and that any overlapping material is protected expression and that it is substantially similar to Plaintiffs' works even after stripping out unprotectable elements like facts, *see Nihon Keizai Shimbun, Inc. v. Comline Business Data*, 166 F.3d 65, 70 (2d Cir. 1999). The user in the referenced conversation provided the News Plaintiff article to Copilot in the prompt. Microsoft further responds that the fact stated in Paragraph 1035 is disputed as the evidence cited in support is currently the subject of a pending motion to exclude expert testimony pursuant to Fed. R. Evid. 702. ECF 2055 at 15-16. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). OpenAI defers to Microsoft's response.

E. Use of News Content in Outputs

1036. "Inference" refers to the process by which the model generates outputs for users. Ex. 800 [Lafferty Rebuttal Rep.] ¶¶ 31-33. *See also* Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶¶ 30-31, 38-40.	1036. **Undisputed.**
1037.Inference begins by feeding prompts into the model, including developer-written system prompts and user prompts. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶¶ 38-40.	1037. **Undisputed**.
1038. A system prompt contains high-level instructions written by the application developer and may instruct the model how to behave and what tools it can use. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 38.	1038. **Undisputed**.
1039. For example, where a product includes a web browser tool, the system prompt can instruct the model that it has a web browser and how and when to trigger browsing in response to a user query. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶ 39.	1039. **Undisputed**.
1040. The LLMs generate outputs by assessing the probability that a given word will follow the previous words in the sequence based on input patterns and outputs reflected across the corpora of training data. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶¶ 30-31; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 138-40.	1040. **Disputed in Part; Not Supported by Cited Evidence.** Defendants respond as follows: Disputed as to the portion of the statement that reads: "based on input patterns and outputs reflected across the corpora of training data" as this is not supported by the cited paragraphs. Rather, LLMs generate outputs based on user inputs, context, a network of up to trillions of

	numerical parameters (mathematical weights) that encode generalized statistical relationships and patterns (such as syntax, semantics, and factual associations), derived from the entire training corpus, that it can use to generate new text using a probability distribution —i.e., to generalize broader ideas and principles of language. *See* ECF 1723 (Lafferty Decl.) ¶¶ 16, 33-35, 45, 52, 56, 64-66, 71-72, 74-77, 80-81, 90-91, 94-100, 115-118, 127-134, 139-41, 154-55, 157-58, 164, 174, 176.
1041. OpenAI's LLMs are autoregressive, meaning that they generate one token at a time by predicting the next token in a text sequence based on the preceding tokens in that sequence. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶¶ 30-31; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 138-139.	1041. **Undisputed.** Microsoft defers to OpenAI.
1042. At each generation step, the model produces a probability distribution across possible next tokens, selects a token, appends that token to the preceding text, and repeats the process until a stop signal. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶¶ 30-31; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 138-139.	1042. **Disputed in Part; Not Supported by Cited Evidence.** Defendants respond as follows: Disputed as to the portion of the statement that reads: "and repeats the process until a stop signal" as that phrase is not supported by the cited paragraphs. Dr. Callison-Burch uses the phrase "this process repeats until the model produces a complete response."
1043. For example, if an LLM receives a sequence like "The cat sat," the pre-trained model will most likely choose "on," but not "a," as the next word. Ex. 461 [Berg-	1043. **Undisputed.** Microsoft defers to OpenAI.

Kirkpatrick Opening Rep.] ¶¶ 30-31; Ex. 483 [Callison-Burch Opening Rep.] ¶¶ 138-139.	
1044. End-users can influence output variability through temperature settings that adjust the next-token probability distribution before selection. Ex. 483 [Callison-Burch Opening Rep.] ¶ 140; Ex. 800 [Lafferty Rebuttal Rep.] ¶ 32.	1044. **Disputed in Part; Not Supported by Cited Evidence.** Defendants respond as follows: Undisputed that the cited evidence supports that temperature is an inference-time parameter that can affect output variability. However, disputed that the cited evidence establishes that end-users control or can adjust that parameter. News Ex. 483 (C. Callison-Burch Opening Rep.) ¶ 140; News Ex. 800 (J. Lafferty Rebuttal Rep.) ¶ 32.
1045. The temperature settings control the "degree of randomness" in the model's selection of the next word. Ex. 483 [Callison-Burch Opening Rep.] ¶ 140.	1045. **Undisputed.** Microsoft defers to OpenAI.
1046. At low temperatures, the model almost always picks the most probable word, while at higher temperatures, it samples more broadly from the distribution. Ex. 483 [Callison-Burch Opening Rep.] ¶ 140.	1046. **Undisputed.** Microsoft defers to OpenAI.
1047.OpenAI has recognized that GPT-4 "can generate plausibly realistic and targeted content, including news articles." Ex. 817 [OPCO_MDL_000157541] at –49.	1047.**Undisputed.** Microsoft defers to OpenAI.
1048. Defendants have produced select subsets of output log data in these cases. [Goldstein Opening Rep.] ¶¶ 382, 421.	1048. **Undisputed**.
1049. Defendants' output log productions did not include conversations that had been deleted after News Plaintiffs filed suit. [Goldstein Opening Rep.] ¶¶ 391, 431-	1049. **Disputed in Part - Not Supported by Cited Evidence.** Defendants respond as follows:

432; Ex. 804 [Berg-Kirkpatrick Rebuttal Rep.] ¶ 268 ("The Merits Sample is best understood as a "retained" sample, because conversations that were subject to a user-initiated deletion are not included.").	Undisputed that the output log productions did not include conversations subject to a user-initiated deletion. However, disputed that the output log production did not include other types of conversations "that had been deleted." The disputed proposition is not supported by the cited evidence. Microsoft further responds that the Copilot output log productions did not include conversations affected by a privacy software program that was running autonomously since before the litigation. ECF 1819-1 (Goldstein Opening Rep.) ¶¶ 391.
1050. Microsoft redacted the approximately 8.2M output logs prior to making them available to News Plaintiffs. In many instances, the redactions were excessive and redacted public information directly from news articles, and nonidentifying information. For example, in one conversation, the following words and phrases are redacted from an article published by the Sun Sentinel: "a lot," "Business Leader of the Year in Broward County," "businessman," "people," "company owner," "mentor," "fundraiser," "board member," "diner patron," and "friend." In other examples, Microsoft has similarly redacted generic and non-identifying terms appearing in public articles. [Goldstein Opening Rep.] ¶ 386.	1050. **Disputed in Part – Not Supported by Admissible Evidence.** Microsoft does not dispute that it redacted the Copilot logs before producing them in order to protect its customers' privacy. But Microsoft disputes that the redactions were excessive or impeded Plaintiffs' investigation. The redactions were targeted at personally identifiable information and other categories independent of Plaintiffs' news content and even included a comprehensive whitelist of terms relating to Plaintiffs including names, organizations, URLs, geographic terms, and email domains. ECF 1723 (Lafferty Decl.) ¶¶ 267-68. Further, plaintiffs' expert Dr. Goldstein devised a methodology explicitly to negate any alleged effect on match counts caused by redactions. *See* ECF 1819-4 (Goldstein Supplemental Rep.) ¶¶ 24-25. Microsoft further responds that the fact stated in Paragraph 1050 is disputed as the evidence cited in support

	is currently the subject of a pending motion to exclude expert testimony pursuant to Fed. R. Evid. 702. ECF 2055 at 17. To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). OpenAI defers to Microsoft's response.
1051. In the set of output logs Microsoft produced, there were over 620 million total redactions. *See* [Goldstein Opening Rep.] ¶ 386; Ex. 800 [Lafferty Rebuttal Rep.] ¶ 134 (identifying 447,124,683 redactions in the "Text" fields and 201,606,256 redactions in the "HiddenText" fields).	1051. **Disputed in Part – Not Supported by Admissible Evidence.** Microsoft does not dispute that output logs Microsoft produced contained over 620 million total redactions. But as stated above, Microsoft disputes that the number of redactions are material to any fact or defense raised in plaintiffs' motion for summary judgment. Microsoft incorporates its response to paragraph 1050 regarding why this paragraph is not supported by admission evidence. Further, only 1.49% of words in user prompts and Copilot outputs were redacted (roughly 447 million redactions out of roughly 30 billion words) and that only 0.60% of words in grounding material were redacted (roughly 202 million redactions out of 39 billion words). Ex. 800 (Lafferty Rebuttal Rep.] ¶¶ 132-134. OpenAI defers to Microsoft's response.
1052. Likewise, OpenAI's output logs contained millions of redactions. Pursuant to a Court order, OpenAI produced an updated version of the 20 million output log sample on January 22, 2026, intended to fix the improper redactions present in the original production. *See*	1052. **Disputed in Part - Improper Legal Argument; Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI reproduced approximately 20 million output logs on January 22, 2026 with amended redactions. ECF 1819-

[Goldstein Opening Rep.] ¶ 440. The updated version still contained over 106,518,208 redactions, or about 5.3 redactions per conversation on average. *See* [Goldstein Opening Rep.] ¶ 449-450.	1 (T. Goldstein Decl. Ex. A (Goldstein Opening Rep.)) ¶ 440. However, disputed as to Plaintiffs' characterization that the reproduction was "intended to fix" any "improper redactions." The Court should disregard the disputed portion of Paragraph 1052 because the assertion constitutes improper attorney argument and/or is conclusory. It thus requires no response. Further, the disputed proposition is not supported by the cited evidence. ECF 1819-1 (T. Goldstein Decl. Ex. A (Goldstein Opening Rep.)) ¶ 440. Microsoft defers to OpenAI.

1. OpenAI's Use of News Content in Outputs

1053.OpenAI's ChatGPT product included News Plaintiffs' content in output shown to a user, including at least substantial portions of News Plaintiffs' Asserted Works.	1053.**Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1053 are disputed because they are not supported by the cited evidence. Plaintiffs do not provide any support for this statement. Microsoft defers to OpenAI.
1054. The messages in which OpenAI's ChatGPT product included The Times's content in output shown to a user include at least the messages from the 20M dataset identified by OpenAI's expert, Dr. Taylor Berg-Kirpatrick. These include the following message ids:	1054. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1054 are disputed because they are not supported by the cited evidence. Specifically, News Plaintiffs' characterization of the cited text matches

b4d3da8b47e0c9babd2d9297f4c77520, 84c7886dfdac7126243e264496970230, fd1446493bae6876bdb0828e8125734c, d395b601298fbf7282a3746c57c282d4, eb306ce7259fa0db2406ac1bc388495b, 7310d53c592d3061b32f8a29cf7d4d83, eab16a41850a50834c2081fa730fedea, b0a5805bba78070cdfd07301102df98c, 3235fae852d32e3c08ecf3aa2ad7f2fc, c1bae243db59cb0146dd08b460dcc3ef, 6cb5bb066cd0f817c9f3887608f3f635. Ex. 461 [Berg-Kirpatrick Opening Rep.] ¶ 384.	identified by Dr. Berg-Kirkpatrick in his analysis of the 20 million conversation logs omits additional context necessary to understand Dr. Berg-Kirkpatrick's findings. Pursuant to Fed. R. Evid. 106, Dr. Berg-Kirkpatrick identified 11 user-visible messages that included a ChatGPT output of a NYT work at issue in this litigation under his methodology. T. Berg-Kirkpatrick Opp. Decl. ¶ 11. Dr. Berg-Kirkpatrick's analysis was not limited to News Plaintiffs' Asserted Works at issue in this Motion. *Id.* at ¶ 12. Further, Plaintiffs have cited no evidence to support the implication that the identified Times content is at issue in this Motion. Microsoft defers to OpenAI.
1055.The messages in which OpenAI's ChatGPT product included Daily News Plaintiffs' content in output shown to a user include at least the messages identified by OpenAI's expert, Dr. Taylor Berg-Kirpatrick. These include the following message ids: 6d938605e08954c5863c822e95138cd1 and 4ccb50e0ff38675fadd6da599e9e0ad3. Ex. 461 [Berg-Kirpatrick Opening Rep.] ¶ 385.	1055. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1055 are disputed because they are not supported by the cited evidence. Specifically, News Plaintiffs' characterization of the cited text matches identified by Dr. Berg-Kirkpatrick in his analysis of the 20 million conversation logs omits additional context necessary to understand Dr. Berg-Kirkpatrick's findings. Pursuant to Fed. R. Evid.106, Dr. Berg-Kirkpatrick did not identify verbatim 25-word outputs in the conversations cited. Rather, he identified two messages containing near-verbatim matches during his robustness check. *See* ECF 1750 (T. Berg-Kirkpatrick Opening Decl.) ¶¶ 32-33.. Dr. Berg-Kirkpatrick's analysis was not limited to News Plaintiffs' Asserted Works at issue in this Motion. T. Berg-

	Kirkpatrick Opp. Decl. ¶¶ 11-12. Further, Plaintiffs have cited no evidence to support the implication that the identified Daily News content is at issue in this Motion. Microsoft defers to OpenAI.
1056. OpenAI's expert did not perform an independent analysis of the 78M and 9M datasets but acknowledged that these datasets include at least 53 conversations, identified by Dr. Goldstein, containing at least a 25-word match to a Times or Daily News Plaintiff Asserted Work. Ex. 802 [Berg-Kirpatrick Supplemental Rep.] ¶ 33.	1056. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1056 are disputed because they are not supported by the cited evidence. Specifically, disputed that Dr. Berg-Kirkpatrick "did not perform an independent analysis of the 78M and 9M datasets." Dr. Berg-Kirkpatrick analyzed their representativeness and examined these datasets. T. Berg-Kirkpatrick Opp. Decl. ¶¶ 17-26. OpenAI further disputes News Plaintiffs' characterization and inferences purportedly drawn therefrom as misleading because it omits necessary context. Pursuant to Fed. R. Evid.106, Dr. Berg-Kirkpatrick did not independently analyze these datasets because they are not representative random samples and overrepresent the prevalence of plugin use and browsing relative to ChatGPT conversations generally. *Id.* Dr. Berg-Kirkpatrick did not extensively search for regurgitation in the 9M and 78M datasets because his assignment was to assess the overall frequency of regurgitation, which requires a representative random sample. *Id.* Dr. Berg-Kirkpatrick's analysis was not limited to News Plaintiffs' Asserted Works at issue in this Motion. *Id.* ¶ 12.

	Further disputed that Dr. Berg-Kirkpatrick "acknowledged that [the 78M and 9M] datasets include at least 53 conversations, identified by Dr. Goldstein, containing at least a 25-word match to a Times or Daily News Plaintiff Asserted Work," as misleading because it omits necessary context. Specifically, Dr. Berg-Kirkpatrick found that the 53 25-word matches related to only 20 unique works and represented 0.00004% of the produced conversation samples. *Id.* ¶ 32. Further, Plaintiffs have cited no evidence to support the implication that the identified Times or Daily News content is at issue in this Motion. Microsoft defers to OpenAI.
1057.The conversations in which OpenAI's ChatGPT product included Ziff Davis content in output shown to a user include at least the conversations identified by OpenAI's expert, Dr. Taylor Berg-Kirpatrick. These include the following conversation id's: f1087e2cd6b524311f934799bfd84034, 810ee325fc98c64741f292acdc4a26c8, 4e691aebd07c52db9eb8ea1f72cae09e, 25123713ac7ff0d9b27b4833309fb698, b44122456df854fabae659a982bcbb29, 5d3171e4551ad108c82809974f64fd59, 2f7ba282895ae3109372cc096256b9f8, b2af680b4f6435a9324e6d15551314d4, 0ceb224345025651e2ff2c2718cd31ff,	1057. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1057 are disputed because they are not supported by the cited evidence. Specifically, News Plaintiffs' characterization of the cited text matches identified by Dr. Berg-Kirkpatrick in his analysis of the 20 million conversation logs omits additional context necessary to understand Dr. Berg-Kirkpatrick's findings. Pursuant to Fed. R. Evid.106, Dr. Berg-Kirkpatrick identified 11 user-visible messages that included a ChatGPT output of a Ziff Davis Plaintiff's work at issue in this litigation under his methodology. ECF 1750 (T. Berg-Kirkpatrick Opening Decl.) ¶ 26, Table 1. As part of his robustness check, Dr. Berg-Kirkpatrick identified 11

0ceb224345025651e2ff2c2718cd31ff, 40b44eced4746e0cc993b3de52fc645a, ce63661f32dc5e7b167facd08708a0fd, 1676f1f6ee6fa025cc7a26719496bc5c, 796ce3742778b16361ff9eb3a47d74ea, 7df6f28432cde03d930daa57345fe49c, 40aadefaf21a8336a657b83e88978652, 93ad118a2e79a34e722abf8bfe6a3a8f, 0a48df489d743a9f9c58ed82018fd10f, 508beb1917cff52c538cf7a6f0a55e16, 955714e397a1bfcf9e2eccc0c488142e. *See e.g.*, Ex. 461 [Berg-Kirpatrick Opening Rep.] ¶¶ 411-412.	additional user-visible messages that included near-verbatim 25-word outputs of a Ziff Davis Plaintiff's works at issue in this litigation. T. Berg-Kirkpatrick Opp. Decl. ¶¶ 53—54. Dr. Berg-Kirkpatrick's analysis was not limited to News Plaintiffs' Asserted Works at issue in this Motion. *Id.* at ¶ 12. The outputs Dr. Berg-Kirkpatrick identified amounted to less than 0.000012% of the 95 million ChatGPT outputs considered from 20 million conversation logs. *Id.* at ¶ 54. . Further, Plaintiffs have cited no evidence to support the implication that the identified Ziff Davis content is at issue in this Motion. Microsoft defers to OpenAI.

i) Regurgitations in OpenAI's Output Logs

1058. OpenAI's logs produced in these cases included numerous examples where ChatGPT regurgitated verbatim text from News Plaintiffs' articles. *See* Goldstein Opening Rep. ¶¶ 366, 462-63, 509; Goldstein Rebuttal Rep. ¶¶ 188-193, 236-237; Goldstein Supplemental Rep. ¶¶ 94-103; Ex. 819 [OPCO_OAI_OUTPUT00009143] at -143, Ex. 821 [OPCO_OAI_OUTPUT00009131] at -131; Ex. 823 [OPCO_OAI_OUTPUT00009183] at -183; Ex. 825 [OPCO_OAI_OUTPUT00006440] at -442; Ex. 827 [OPCO_OAI_OUTPUT00006740] at 740-42; Ex. 831 [OPCO_OAI_OUTPUT00006563] at –563-64; Ex. 1261	1058. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1058 are disputed because they are not supported by the cited evidence. Specifically, News Plaintiffs' claim of "numerous examples" is vague, misleading, and contradicted by record evidence. For example, the combined regurgitation, summarization, and other claimed content reproduction examples claimed by NYT and Daily News Plaintiff expert Dr. Goldstein across his Opening, Rebuttal, and Supplemental Reports amount to less than 0.00011% of conversations out of the 118M produced examples. T. Berg-Kirkpatrick Opp. Decl. ¶ 46.

[OPCO_OAI_OUTPUT00006728] at –-28-29; Ex. 833 [OPCO_OAI_OUTPUT00006515] at -518; Ex. 835 [OPCO_OAI_OUTPUT00006565] at -565-66.	Microsoft defers to OpenAI.
1059. OpenAI's technical expert Dr. Berg-Kirkpatrick found millions of verbatim 25-word spans of text ("25-grams") contained in News Plaintiffs' asserted works throughout the 20 million ChatGPT conversations: The Times: 409,219 instances of 25-grams spread across 19,330 unique conversations drawn from 3,191 unique articles; *Daily News* Plaintiffs: 826,660 instances of 25-grams spread across 31,500 unique conversations drawn from 2,365 unique articles; CIR: 406,182 instances of 25-grams spread across 14,285 unique conversations drawn from 207 unique articles; The Intercept: 38,293 instances of 25-grams spread across 1,583 unique conversations drawn from 166 unique articles; *Ziff Davis* Plaintiffs: 5,296,745 instances of 25-grams spread across 39,115 unique conversations drawn from 3,622 unique articles. Ex. 461 [Berg-Kirkpatrick Opening Rep.] ¶¶ 210, 387, 394, 400, 407).	1059. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1059 are disputed because they are not supported by the cited evidence. *See* News Ex. 461 (T. Berg-Kirkpatrick Opening Rep.) ¶¶ 210, 387, 394, 400, 407. Specifically, the fact omits context necessary to understand Dr. Berg-Kirkpatrick's analysis and mischaracterizes his findings. News Plaintiffs cite Dr. Berg-Kirkpatrick's figures for "25-grams" identified before matches to boilerplate sequences, third-party material, and text contained within the user's prompt were removed. ECF 1750 (T. Berg-Kirkpatrick Opening Decl.) ¶¶ 17-20; T. Berg-Kirkpatrick Opp. Decl. ¶¶ 13-14. These "25-grams" included text matches to, for example, NYT works at issue in this litigation that are transcripts of speeches and legal documents *Id.* at ¶ 14, n. 24.. These figures do not reflect instances of 25-word excerpts of Plaintiff-generated text appearing in user-visible ChatGPT outputs. *Id.* Further disputed that these figures represent matches "contained in News Plaintiffs' asserted works." Pursuant to Fed. R. Evid.106, these figures were computed before News Plaintiffs made significant concessions removing works from their asserted works corpora. ECF 1819-3 (T. Goldstein Decl. Ex. C (Reply Rep.)) ¶¶ 68-69, 196-205 (updating NYT/DNP Rule 1006 collections); ECF 1774-1

	(D. Delorey Decl. Ex. A (Opening Rep.)) ¶ 140 (removing nearly 200,000 Asserted Works). Dr. Berg-Kirkpatrick's analysis was not limited to News Plaintiffs' Asserted Works at issue in this Motion. T. Berg-Kirkpatrick Opp. Decl. ¶ 12. Plaintiffs have cited no evidence to support the implication that the identified News Plaintiffs' content is at issue in this Motion. Microsoft defers to OpenAI.

a) ChatGPT 20M Sample (sample_conversations)

1060. In the conversation (id: 810ee325fc98c64741f292acdc4a26c8) reproduced below, ChatGPT reproduced portions of a Lifehacker article "A Beginner's Guide to Using ChatGPT Plugins." Ex. 461 [Berg-Kirkpatrick Opening Report] at 197.	1060. **Disputed – Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1060 are disputed because they are not supported by the cited evidence. *See* News Ex. 461 (T. Berg-Kirkpatrick Opening Rep.) at p. 197. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "reproduce[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' work at issue in this litigation appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are

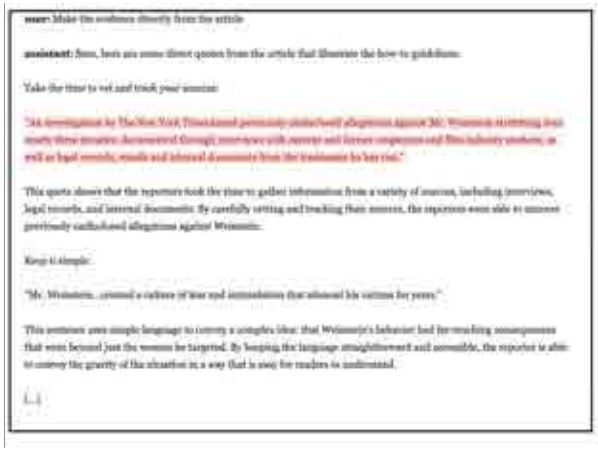
	so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid. 106, the cited message included matched text of only a small fraction of the Asserted Work at issue. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 24, 63. Microsoft defers to OpenAI.

1061. In the conversation (id: f1087e2cd6b524311f934799bfd84034) reproduced below, ChatGPT reproduced Eurogamer article "Twitch introducing Stories and Discovery Feed to boost offline community interaction." Ex. 461 (Berg-Kirkpatrick Opening Report) at 196. 	1061. **Disputed – Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1061 are disputed because they are not supported by the cited evidence. *See* News Ex. 461 (T. Berg-Kirkpatrick Opening Rep.) at 196. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "reproduce[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' works at issue in this litigation appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid. 106, the cited message included matched text of only a small fraction of the Asserted Work at issue. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 24, 63. Microsoft defers to OpenAI.

b) ChatGPT 78M Sample (giraffe_chatgpt)

1062. In another conversation (id: caaee9ab-4094-44f7-8d75-a5ae69e45ee6), reproduced below, ChatGPT provides 45 words from the Times article titled "Public &	1062. **Disputed.** OpenAI responds as follows: The facts set forth in Paragraph 1062 are disputed because they are not supported by the cited evidence. Specifically,

Private; Dumb About Smarts." The red text represents the text from ChatGPT's output that overlaps with the original article. Goldstein Rebuttal Rep. ¶ 190; Ex. 819 [OPCO_OAI_OUTPUT00009143] at -143.  **user:** best Future quotes **assistant:** Here are a few potential quotes about the future: - "The future belongs to those who believe in the beauty of their dreams." - Eleanor Roosevelt - "The best way to predict the future is to create it." - Peter Drucker - "The future is something which everyone reaches at the rate of sixty minutes an hour, whatever he does, whatever he is." - C.S. Lewis - "The future is not a gift. It is an achievement." - Robert F. Kennedy - "The future is not some place we are going to, but one we are creating. The paths to it are not found but made, and the activity of making them changes both the maker and the destination." - John Schaar - "The future is not set, there is no fate but what we make for ourselves." - Sarah Connor, Terminator 2: Judgment Day - "The future depends on what we do in the present." - Mahatma Gandhi - "The future is built on brains, not prom court, as most people can tell you after attending their high school reunion. But you'll never know it by talking to kids or listening to the messages they get from the culture and even from their schools." - Susan Cain - "The future is an opaque mirror. Anyone who tries to look into it sees nothing but the dim outlines of an old and worried face." - Jim Bishop - "The future is always beginning now." - Mark Strand	OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "provide[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' works at issue in this litigation appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid.106, the cited message included matched text of only a small fraction of the Asserted Work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.
1063. In another conversation (id: b366fe02-b855-42ff-a207-26b2ed1a2150), reproduced below, ChatGPT output 83 words from the Times article titled "Harvey Weinstein Paid Off Sexual Harassment Accusers for Decades." Goldstein Rebuttal Rep ¶ 191; Ex. 821 [OPCO_OAI_OUTPUT00009131] at -131.	1063. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The facts stated in Paragraph 1063 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not a Times Asserted Work at issue in this Motion.

 	Further, the facts set forth in Paragraph 1063 are disputed because they are not supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "output[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' works at issue appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are so rare that the vast majority of ChatGPT users will never encounter one. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 33. Further, pursuant to Fed. R. Evid.106, the cited message included matched text of only a small fraction of a work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.
1064. In another conversation (id: f9c9f848-ab5d-490b-bf4e-32f91c8a8d9d), reproduced below, ChatGPT output 34 words from the Times article titled "The Moral	1064. **Disputed – Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1064 are disputed because they are not

Bucket List." The red text represents the text from ChatGPT's output that overlaps with the original article. Goldstein Rebuttal Rep. ¶ 192; Ex. 823 [OPCO_OAI_OUTPUT00009183] at -183.	supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "output[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' works at issue in this litigation appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are so rare that the vast majority of ChatGPT users will never encounter one. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 33.Further, pursuant to Fed. R. Evid. 106, the cited message included matched text of only a small fraction of the Asserted Work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.
1065. In the conversation (id: 1698f8ee-71fc-4ded-8b64-b5246b3f244b) reproduced below, ChatGPT reproduces the verbatim text from the Times article titled "Screen: 'Bonnie and Clyde' Arrives." The red text represents the text from ChatGPT's output that overlaps with the original article. Ex. 804 [Berg-Kirkpatrick Rebuttal Rep.] ¶ 189.	1065. **Disputed – Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1065 are disputed because they are not supported by the cited evidence. *See* News Ex. 804 (T. Berg-Kirkpatrick Rebuttal Rep.) ¶ 189. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "reproduces"

Table 26: Example #1 of "Regurgitation" Conversation 	Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' works at issue in this litigation appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid. 106, the cited message included matched text of only a small fraction of the Asserted Work *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.

c) ChatGPT 9M Sample (chatgpt_20230714)

1066. In the conversation (id: 2ec49781-3263-46af-bf81-0761260fe158) reproduced below, ChatGPT retrieved over 600 words from the Times article titled "Cracking Down on Dissent, Russia Seeds a Surveillance Supply Chain" and proceeded to provide over 200 words of that article in an output to a user. The red text represents the text from ChatGPT's output that overlaps with the original article. Goldstein Supplemental Rep. ¶ 95; Ex. 825 [OPCO_OAI_OUTPUT00006440] at -442.	1066. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The facts stated in Paragraph 1066 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not a Times Asserted Work at issue in this Motion. Further, the facts set forth in Paragraph 1066 are disputed because they are not supported by the cited evidence. Specifically, the cited conversation is drawn from the 9M dataset, which is not a representative random sample and overrepresents the prevalence of plugin use and browsing relative to ChatGPT conversations generally. Further disputed because News Plaintiffs' characterization of the cited conversation, and inferences drawn therefrom, is misleading. Pursuant to Fed. R. Evid. 106, this conversation involved the use of a third-party tool. Third-party tools are not part of OpenAI's native Browse/Search functionality, nor are they operated by OpenAI. These third-party tools were not created by OpenAI, and OpenAI's native tools associated with the Browse function were not performing the search or actively visiting the Plaintiffs' domains in these cases. *See* ECF 1756 (N. Feamster Decl.) ¶¶ 26–27. Further, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed

	95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Pursuant to Fed. R. Evid. 106, the cited message included matched text of only a small fraction of the Asserted Work: approximately 200 out of approximately 1,800 words. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.
1067.In the conversation (id: cecb4afb-7607-462c-b4eb-9e19dd6d989e) reproduced below, ChatGPT retrieved nearly 100 words from the Mercury News article titled "How many of California's homeless residents are	1067. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The facts stated in Paragraph 1067 are immaterial and therefore disputed because these facts do not affect the outcome of

from out of state?" and proceeded to provide over 70 words from the article in an output to a user. The red text represents the text from ChatGPT's output that overlaps with the original article. Goldstein Supplemental Rep. ¶ 96; Ex. 827 [OPCO_OAI_OUTPUT00006740] at -740-741.



the suit under the governing law. Specifically, the cited article is not a Daily News Asserted Work at issue in this Motion.

Further, the facts set forth in Paragraph 1067 are disputed because they are not supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33.

Further, pursuant to Fed. R. Evid.106, the cited message included matched text of only a small fraction of the Asserted Work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63.

Microsoft defers to OpenAI.

1068. In the conversation (id: cecb4afb-7607-462c-b4eb-9e19dd6d989e) reproduced below, ChatGPT retrieved nearly 200 words from the Mercury News article titled "DeSantis releases ad blasting 'madness' of 'once-great' San Francisco" and proceeded to output the entirety of the retrieved article text in its output to the user. The red text represents the text from ChatGPT's output that overlaps with the original article. Goldstein Supplemental Rep. ¶ 97; Ex. 827 [OPCO_OAI_OUTPUT00006740] at -741.	1068. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The facts stated in Paragraph 1068 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not a Daily News Asserted Work at issue in this Motion. Further, the facts set forth in Paragraph 1068 are disputed because they are not supported by the cited evidence.. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid.106, the cited message included matched text of only a small fraction of the Asserted Work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.

	
1069. In the conversation (id: cecb4afb-7607-462c-b4eb-9e19dd6d989e) reproduced below, ChatGPT retrieved over 200 words from the Mercury News article titled "Battle erupts over homes to replace San Jose swim club and tennis site" and proceeded to output a summary containing verbatim snippets from the article. The red text represents the text from ChatGPT's output that overlaps with the original article. Goldstein Supplemental Rep. ¶	1069. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The facts stated in Paragraph 1069 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not a Daily News Asserted Work at issue in this Motion. Further, the facts set forth in Paragraph 1069 are disputed because they are not supported by the cited

98; Ex. 827 [OPCO_OAI_OUTPUT00006740] at -741-742. 	evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid.106, the cited message included matched text in red of only a small fraction of the Asserted Work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.
1070. In the conversation (id: 97fde30c-c108-404b-ad64-23c00b9481c9) reproduced below, ChatGPT retrieved over 600 words from the Times article titled "Turkey Clears the Way for Sweden's Entry to NATO on the Eve of Summit" and proceeded to provide nearly 80 words of that article in an output to a user. The red text represents the text from ChatGPT's output that overlaps with the	1070. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The facts stated in Paragraph 1070 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not a Times Asserted Work at issue in this Motion.

original article. Goldstein Supplemental Rep. ¶ 99; Ex. 831 [OPCO_OAI_OUTPUT00006563] at –563-64. 	Further, the facts set forth in Paragraph 1070 are disputed because they are not supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" News Plaintiffs' content in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs to identify contiguous sequences of 25 words of output text that matched Plaintiffs' article text. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on Dr. Berg-Kirkpatrick's analysis, this output does not contain any 25-word excerpts of the Plaintiffs' Asserted Works. Dr. Berg-Kirkpatrick found that 25-word verbatim excerpts of News Plaintiffs' works at issue in this litigation appeared in less than 0.000025% of the analyzed outputs, demonstrating that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are extremely rare. *Id.* ¶ 33. Microsoft defers to OpenAI.
1071. In the conversation (id: a01433eb-a1ef-4d22-bbe8-70e4b02f50f9) reproduced below, ChatGPT	1071. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The

retrieved of nearly 650 words from the Times article titled "Trump and Allies Forge Plans to Increase Presidential Power in 2025" and proceeded to provide nearly 190 words from the article in an output to the user. The red text represents the text from ChatGPT's output that overlaps with the original article. Ex. 1261 [OPCO_OAI_OUTPUT00006728] at –-28-29.	facts stated in Paragraph 1071 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not a Times Asserted Work at issue in this Motion. Further, the facts set forth in Paragraph 1071 are disputed because they are not supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid. 106, the cited message included matched text in red of only a small fraction of the Asserted Work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.

	
1072. In the conversation (id: 86734e5a-809b-4eb4-81df-3b6f3bd5cb22) reproduced below, ChatGPT	1072. **Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as

retrieved over 650 words from the Times article titled "For Privacy From Cranks, Creeps and Crooks, Millions Getting Unlisted Phone Numbers" and proceeded to output a summary having nearly 50 words overlapping with the text of that article. The red text represents the text from ChatGPT's output that overlaps with the original article. Ex. 837 [OPCO_OAI_OUTPUT00006651] at -652-53.	follows: The facts stated in Paragraph 1072 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Specifically, the cited article is not a Times Asserted Work at issue in this Motion. Further, the facts set forth in Paragraph 1072 are disputed because they are not supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" News Plaintiffs' content in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs to identify contiguous sequences of 25 words of output text that matched Plaintiffs' article text. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on Dr. Berg-Kirkpatrick's analysis, this output does not contain any 25-word excerpts of the Plaintiffs' works at issue in this litigation. Dr. Berg-Kirkpatrick found that 25-word verbatim excerpts of News Plaintiffs' works at issue in this litigation appeared in less than 0.000025% of the analyzed outputs, demonstrating that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are extremely rare. *Id.* ¶ 33. Microsoft defers to OpenAI.

	
1073. In the conversation (id: 8e875b57-f9ea-4a94-83d2-34182c51c191) reproduced below, ChatGPT retrieved over 500 words from the Times article titled "The	1073. **Disputed in Part– Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1073 are disputed in part because

Southwest Bakes With More Heat on the Way" and output more than 100 words from that article to the user. The red text represents the text from ChatGPT's output that overlaps with the original article. Goldstein Supplemental Rep. ¶ 100; Ex. 833 [OPCO_OAI_OUTPUT00006515] at -518. 	they are not supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" Plaintiffs' Asserted Works in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs and found that 25-word verbatim excerpts of News Plaintiffs' Asserted Works appeared in less than 0.000025% of the analyzed outputs. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on this analysis, Dr. Berg-Kirkpatrick concluded that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' Asserted Works are so rare that the vast majority of ChatGPT users will never encounter one. *Id.* ¶ 33. Further, pursuant to Fed. R. Evid. 106, the cited message included matched text in red of only a small fraction of the Asserted Work. *See* ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶ 24, 63. Microsoft defers to OpenAI.
1074. In the conversation (id: 98c2eccd-546b-4592-bfc5-8ca6e4a30a26) reproduced below, ChatGPT retrieved over 740 words from the Times article titled	1074. **Disputed in Part– Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1074 are disputed in part because

"Threads Becomes Most Rapidly Downloaded App, Raising Twitter's Ire" and proceeded to provide nearly 140 words of the article in an output to a user. The red text represents the text from ChatGPT's output that overlaps with the original article. Ex. 835 [OPCO_OAI_OUTPUT00006565] at -565-66.	they are not supported by the cited evidence. Specifically, OpenAI disputes the implication that the cited user-visible message is evidence that ChatGPT "retrieve[s]" News Plaintiffs' content in everyday conversations with users. Dr. Berg-Kirkpatrick analyzed 95 million user-visible ChatGPT outputs to identify contiguous sequences of 25 words of output text that matched Plaintiffs' article text. ECF 1750 (T. Berg-Kirkpatrick Decl.) ¶¶ 10–27. Based on Dr. Berg-Kirkpatrick's analysis, this output does not contain any 25-word excerpts of the Plaintiffs' works at issue in this litigation. Dr. Berg-Kirkpatrick found that 25-word verbatim excerpts of News Plaintiffs' works at issue in this litigation appeared in less than 0.000025% of the analyzed outputs, demonstrating that ChatGPT outputs containing verbatim or near-verbatim portions of Plaintiffs' works at issue in this litigation are extremely rare. *Id.* ¶ 33. Microsoft defers to OpenAI.



ii) Other Uses of News Plaintiffs' Content in OpenAI's Output Logs

1075. In addition to verbatim reproductions of News Plaintiffs' content, Defendants' output logs contain other types of substitutive and harmful outputs that do not contain verbatim copies of News Plaintiffs' content but are the result of Defendants having trained and/or grounded on copies of News Plaintiffs' content, including summaries of News Plaintiffs' articles, content generated in the style of News Plaintiffs' publications, and rewritten News Plaintiffs' articles. Goldstein Opening Rep. ¶¶ 464-515; Goldstein Rebuttal Rep. ¶¶ 195-212; Goldstein Supplemental Rep. ¶¶ 103-116.	1075. **Disputed - Not Supported by Cited Evidence.** Defendants respond as follows: The facts set forth in Paragraph 1075 are disputed because they are not supported by the cited evidence. Specifically, disputed that Defendants' output logs contain "substitutive and harmful outputs." Disputed that any particular output was "the result of Defendants having trained and/or grounded on copies of News Plaintiffs' content." Disputed that "summaries of News Plaintiffs' articles, content generated in the style of News Plaintiffs' publications, and rewritten News Plaintiffs' articles" are "substitutive and harmful outputs." *See* Lafferty Decl. ¶¶ 21, 28 (Table 1 (Rebuttal), 175, 187, 331, 398, 408-10; Tucker Decl. ¶¶ 56-89, 92-104 Defendants further dispute Paragraph 1075 as stating a legal conclusion to the extent it suggests that alleged outputs that summarize, are allegedly "in the style of," or are "rewritten" Asserted Works are unlawful under the meaning of copyright law; the Court should disregard this disputed portion of Paragraph 1075 because the assertion constitutes improper attorney argument and is conclusory. It thus requires no response.
1076. Many of the ChatGPT conversation records produced to Plaintiffs replaced the URL from the actual conversation with "<URL>." *See, e.g., Ex. 881* [OPCO_OAI_OUTPUT_00009067] at -067.	1076. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1076 are not supported by the cited evidence. Specifically, disputed as to Plaintiffs' use of "many" as vague, ambiguous, and unsupported by the cited

	evidence. News Ex. 881 only shows 5 instances of "." Further disputed to the extent Plaintiffs suggest any URL replacements were applied as part of production to Plaintiffs. Pursuant to Fed. R. Evid.106, the conversation Plaintiffs cite in support is from the 78M dataset which Plaintiffs knew had been deidentified prior to this litigation. Ex. 240 (V. Monaco 30(b)(6) Tr. Apr. 8, 2026) 245:22-23; ECF 1312 at 3 (Plaintiffs requested "[a]ll of OpenAI's reservoirs of de-identified and searchable output logs, including but not limited to the 10 million and 78 million reservoirs"). Microsoft defers to OpenAI.
1077. Notwithstanding OpenAI's redactions to its produced output logs, OpenAI's expert, Dr. Berg-Kirkpatrick, identified tens of thousands of user queries in the 20 million conversation dataset in which the user expressly referenced News Plaintiffs by name when prompting ChatGPT. Ex. 461 [Berg-Kirkpatrick Opening Rep.] at 252-53 (Table A.48).	1077. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that OpenAI's expert, Dr. Berg-Kirkpatrick, used the keywords cited in Table A.48 to "identif[y] all user requests in the entire Merits Sample whose text included the name of a News Plaintiff." News Ex. 461 (T. Berg-Kirkpatrick Opening Rep.) ¶ 177; *id.* at 252–53 (Table A.48). However, disputed as irrelevant and immaterial to the resolution of this motion that a user referenced a keyword used by Dr. Berg-Kirkpatrick in a prompt. The facts stated in Paragraph 1077 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.

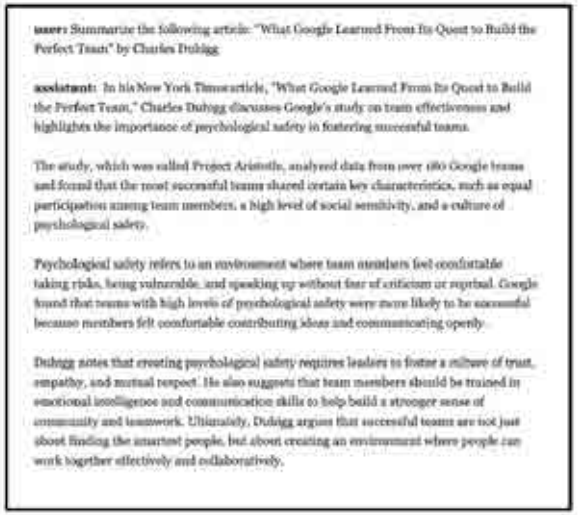
a) ChatGPT 20M Sample (sample_conversations)

1078. In the conversation (id: 75121d54f1bb0ba46ab0946e05aa9252) the user asked ChatGPT to "give me a summary of The Playbook for Poisoning the Earth that is written by the intercept." ChatGPT responded with the following. Ex. 839 [OPCO_OAI_OUTPUT00004962] at -962, -967.	1078. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that the cited conversation includes the stated prompt and output. However, the remainder of Paragraph 1078 is disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), Dkt. No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."); Fed. R. Evid. 106. Therefore, the disputed proposition in Paragraph 1078 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1079. In the conversation (id: 56dce8c6ab4ba080aa476ed70fd63409) the user provided the author, title, and URL of a Mother Jones article and asked, "can you please write a small paragraph summarising this article." ChatGPT responded, "Sure, I can provide a summary based on the information available in the article:\n\nIn her article \"Charts: The State of Women\u2019s Athletics, 40 Years After Title IX." It then	1079. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that the cited conversation includes the stated prompt and output. However, disputed as misleading due to the omission of necessary context. Pursuant to Fed. R. Evid. 106, earlier in the conversation, ChatGPT responded to a related user prompt as follows: "I'm sorry for any confusion, but I can't provide direct links to specific articles, especially those behind paywalls or hosted on

provided the following summary. Ex. 871 [OPCO_OAI_OUTPUT00004276] -283-84. 	proprietary databases, as I don't have direct access to external content. However, you can typically find the article by searching for the title (\"Gender Pay Gap in Sports: A Review of Literature and Future Research Agenda\") in academic databases like PubMed, Google Scholar, or by accessing it through your university's library resources. If you have any other questions or need further assistance, feel free to ask!" News Ex. 871 at OPCO_OAI_OUTPUT000004277. Further disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), Dkt. No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1079 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.

b) ChatGPT 78M Sample (giraffe_chatgpt)

1080. In one conversation (id: c56bd90a-4cc4-463f-ae79-03f1f38a61f4), the user asked ChatGPT to "summarize the following article" and provided the title and author of a Times article. ChatGPT responded by providing a four-	1080. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that the cited conversation includes the stated prompt and output.

paragraph summary of the article. *See* [Goldstein Rebuttal Rep.] ¶ 208; Ex. 873 [OPCO_OAI_OUTPUT00009137] at -137.  **user:** Summarize the following article: "What Google Learned From Its Quest to Build the Perfect Team" by Charles Duhigg **assistant:** In his New York Times article, "What Google Learned From Its Quest to Build the Perfect Team," Charles Duhigg discusses Google's study on team effectiveness and highlights the importance of psychological safety in fostering successful teams. The study, which was called Project Aristotle, analyzed data from over 180 Google teams and found that the most successful teams shared certain key characteristics, such as equal participation among team members, a high level of social sensitivity, and a culture of psychological safety. Psychological safety refers to an environment where team members feel comfortable taking risks, being vulnerable, and speaking up without fear of criticism or reprisal. Google found that teams with high levels of psychological safety were more likely to be successful because members felt comfortable contributing ideas and communicating openly. Duhigg notes that creating psychological safety requires leaders to foster a culture of trust, empathy, and mutual respect. He also suggests that team members should be trained in emotional intelligence and communication skills to help build a stronger sense of community and teamwork. Ultimately, Duhigg argues that successful teams are not just about finding the smartest people, but about creating an environment where people can work together effectively and collaboratively.	However, disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1080 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1081. In another conversation (id: f60dc43e-b500-4d0b-99fd-3c022d36d7b7), the user asked ChatGPT for a "summary" of a Times article, identified by title, date and exact URL (redacted by OpenAI), and ChatGPT provided a four-paragraph summary of the article. *See* [Goldstein Rebuttal Rep.] ¶ 209; Ex. 875 [OPCO_OAI_OUTPUT00009185] at -185.	1081. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that the cited conversation includes the stated prompt and output. However, disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1081 is immaterial because it does not affect the outcome of the suit under the governing law.

 **user:** give me a 250 word summary on the following article Plumer, Brad, and Nadja Popovich. "How Decades of Racist Housing Policy Left Neighborhoods Sweltering." NewYorkTimes.com, 24 August 2020, <URL> **assistant:** The article "How Decades of Racist Housing Policy Left Neighborhoods Sweltering" published in The New York Times explores the connection between historical redlining practices and the disproportionate impact of heatwaves on communities of color in American cities today. Redlining refers to a discriminatory practice that prevented Black families from accessing loans, insurance, and other services in certain neighborhoods. The practice began in the 1930s and was formally outlawed in 1968, but its legacy is still felt in American cities today. The article explains that the areas that were redlined in the past are now the same areas that experience the highest temperatures during heatwaves, which are becoming more frequent and severe due to climate change. The article provides examples from cities like Baltimore, St. Louis, and New York, where neighborhoods that were historically redlined have fewer trees, parks, and green spaces that can help mitigate the effects of extreme heat. These areas also have a higher concentration of heat-absorbing surfaces like asphalt and concrete. As a result, residents in these neighborhoods are more likely to suffer from heat-related illnesses and even death during heatwaves. The article argues that the current heat crisis is not only a result of climate change but also of systemic racism that has prevented Black communities from accessing the resources they need to adapt to a changing climate. The article concludes by calling for policy changes that address the root causes of this problem, including investment in green infrastructure, affordable housing, and equitable access to services for all communities.	Microsoft defers to OpenAI.
1082. In another conversation, a user asked ChatGPT "Tell me about the CIA selling crack to the inner cities in the 1980s and '90s." In response, ChatGPT provided a detailed summary of the Mercury News reporting on this topic. Goldstein Rebuttal Rep. ¶ 212; Ex. 877 [OPCO_OAI_OUTPUT00009075] at -075.	1082. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed as to the content of the user prompt. However, disputed as to the characterization that ChatGPT provided a "detailed" summary of the Mercury News reporting on this topic. *Cf.* News Ex. 877 at OPCO_OAI_OUTPUT00009075. The assistant response to the user prompt was a total of 364 words, which News Plaintiffs do not assert contained any verbatim passages of the relevant articles. Approximately 40% of the assistant response addressed relevant issues that were not covered by Gary Webb's reporting in the Mercury News. Further disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's

	LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1082 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1083. In another example (id: 716ce74f-d821-4b70-8452-f233ca0b1b48), in response to a user query that included the title of the Times article, "6 Black Chefs (and 1 Inventor) Who Changed the History of Food," ChatGPT responded by providing a seven-paragraph summary of the article. *See* [Goldstein Rebuttal Rep]. ¶ 205; Ex. 879 [OPCO_OAI_OUTPUT_00009090] at -090.	1083. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed as to the content of the user prompt. However, disputed as to the characterization of the assistant response, which was a 393-word summary that News Plaintiffs do not assert included any verbatim passages from the relevant article. Further disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying

	facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1083 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1084. In one conversation (id: 62f33cb1-a58a-4270-b491-adfb4247e157), a user submitted multiple queries asking ChatGPT to provide summaries of news articles and provided URLs to each article. Two of these queries requested summaries of news articles from the New York Times and the Denver Post. Goldstein Rebuttal Rep. ¶ 211; Ex. 881 [OPCO_OAI_OUTPUT_00009067] at -067.	1084. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed that the cited conversation includes the stated prompt and output. However, disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1084 is immaterial because it does not affect the outcome of the suit under the governing law.



user: summarize <URL>

assistant: The New York Times article reports on a harm reduction program in Iowa that has been successful in reducing overdose deaths. The program involves the distribution of naloxone, a medication that can reverse opioid overdoses, as well as fentanyl test strips to detect the presence of the deadly synthetic drug in street drugs.

According to the article, the program has been instrumental in saving lives, with a significant decrease in overdose deaths in areas where it has been implemented. The article highlights the success of the program in contrast to traditional drug policies that focus on punishment and abstinence.

The article also notes some opposition to the harm reduction approach, with some critics arguing that it enables drug use and may lead to an increase in drug-related crime. However, advocates argue that harm reduction programs are a more realistic and effective way of addressing the ongoing drug overdose crisis.

Overall, the article provides an informative account of the harm reduction program in Iowa, highlighting its success in reducing overdose deaths and its potential as a model for addressing the ongoing drug overdose crisis in the United States.



user: summarize <URL>

assistant: The Denver Post article reports on a proposed bill in the Colorado legislature aimed at preventing drug overdose deaths, particularly those related to the synthetic opioid fentanyl. The bill would require the state's public health department to develop a program for distributing free fentanyl test strips and naloxone to individuals and organizations, as well as provide funding for community-based overdose prevention programs.

The article notes that fentanyl-related deaths have been on the rise in Colorado, mirroring a trend seen across the country. The proposed bill seeks to address this crisis by providing individuals with the tools and resources needed to reduce the risk of overdose and save lives.

The article also discusses some of the opposition to the bill, with some critics arguing that it could encourage drug use and send the wrong message to young people. However, supporters of the bill argue that it is a necessary and evidence-based approach to addressing the ongoing overdose crisis.

Overall, the article provides a detailed account of the proposed bill in the Colorado legislature aimed at preventing drug overdose deaths, highlighting its potential impact on reducing overdose deaths and its potential challenges and opposition.

Microsoft defers to OpenAI.

1085. In another conversation (id: d991d506-d52e-4caf-b6a5-7f5dc4439014), reproduced below, a user asked ChatGPT to "summarize the article 'New Room-Temperature Superconductor Offers Tantalizing Possibilities' from New York Times" and ChatGPT responded with a one-paragraph summary. The user followed up asking ChatGPT to "summarize in more details" and ChatGPT responded with "a more detailed summary." Ex. 883 [OPCO_OAI_OUTPUT00009156] at -156. 	1085. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed as to the content of the user prompt. However, disputed as to the characterization of the assistant responses. The initial assistant response was a 112-word summary. The follow-up response that News Plaintiffs characterize as a "more detailed" summary was 250 words. News Plaintiffs do not contend either response contained any verbatim passages from the article identified in the user prompt. Further disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1085 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1086. In another conversation (id: 12e25881-3d47-49a2-afbd-7676f66e3c1e), reproduced below, a user provides ChatGPT with a URL to a Times article, requesting a summary, and ChatGPT provided a detailed	1086. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed as to the content of the user prompt.

summary of the article. Ex. 885 [OPCO_OAI_OUTPUT00009045] at -045. 	However, disputed as to the characterization of the assistant response as "detailed." The assistant response was 148 words, which News Plaintiffs do not assert contained any verbatim passages from the article identified in the user prompt. Further disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1086 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1087. In the conversation (id: ed05384c-0e81-40f5-b01c-c83c3a38281e) reproduced below, a user asked ChatGPT to provide an "example of article about current news wit[h] source" and ChatGPT returned a summary of a Times article. Ex. 887 [OPCO_OAI_OUTPUT_00009169] at -169.	1087. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed as to the content of the user prompt. However, disputed as to the characterization of the assistant response, which was 310 words, with 57 of those

	words quotes from Joe Biden and Gina McCarthy, White House national climate advisor. Further disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op. at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1087 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1088. In another conversation (id: 74cb570f-da55-46e8-8917-eb9e8054fbfb), a user asked ChatGPT to summarize a Chicago Tribune article and ChatGPT returned a one paragraph summary. When the user followed up by stating "Give me more," ChatGPT provided 4 additional paragraphs summarizing the article. Ex. 889 [OPCO_OAI_OUTPUT_00009071] at -071.	1088. **Disputed in Part - Immaterial.** OpenAI responds as follows: Undisputed as to the content of the user prompt. However, disputed as to the characterization of the assistant responses. The initial assistant response was a 109-word summary. The follow-up response that News Plaintiffs characterize as providing "4 additional paragraphs" was 214 words. News Plaintiffs do not contend either response contained any verbatim passages from the article in question. Further disputed as immaterial. This Court has already rejected News Plaintiffs' claim that OpenAI's LLMs infringe their works by summarizing the facts in them. Op.

	at 41, *New York Times v. Microsoft*, No. 1:23-cv-11195-SHS-OTW (S.D.N.Y. Apr. 4, 2025), ECF No. 514 ("Works are not substantially similar as a matter of law when their only similarity concerns general or underlying facts and ideas, which are not copyrightable."). Therefore, the disputed proposition in Paragraph 1088 is immaterial because it does not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.

2. Microsoft's Use of News Content in Outputs

1089. In one conversation (51D|BingProd|74F0DD2A7CF9E3794B5C61D396D519E0A8FD756EDE58D55F6442AFFFA87B7C6B) Copilot retrieved over 500 words from the Chicago Tribune article "The 1979 Crash of Flight 191 Remembered in Editorial from 2004" and proceeded to provide the user with a paraphrased version of the article. The text retrieved by Copilot is reproduced below. The red text represents the text from the RAG data retrieved by Copilot that overlaps with	1089. **Disputed – Not Supported by Cited Evidence; Improper Legal Argument; Not Supported by Admissible Evidence.** Microsoft responds that the facts set forth in Paragraph 1089 are disputed because the cited evidence does not support the statement. Plaintiffs cite News Ex. 806 at MSFT_AI_MDL_002882780-783. This text does not appear at that pincite, and the exhibit contains no red

the original article. Ex. 806 [MSFT_AI_MDL_002882532] at -780-783. 	lettering; Plaintiffs have not identified where this screenshot—or the red lettering within it—comes from. Microsoft further responds that this fact constitutes an improper legal conclusion, as it presumes the volitional conduct requirement for direct infringement, *see Cartoon Network*, 536 F.3d at 131, and that any overlapping material is protected expression and that it is substantially similar to Plaintiffs' works even after stripping out unprotectable elements like facts, *see Nihon Keizai Shimbun, Inc. v. Comline Business Data*, 166 F.3d 65, 70 (2d Cir. 1999). Microsoft further responds that Paragraph 1089 is disputed as the evidence cited in support is currently the subject of a pending motion to exclude expert testimony pursuant to Fed. R. Evid. 702. ECF No. 2055 at II.A.1 & II.A.3 (challenging the methodology used to find "matches"). To the extent the Court excludes the challenged opinions, Plaintiffs may not rely on them to support this statement. *See* Fed. R. Civ. P. 56(c)(2). OpenAI defers to Microsoft's response.
1090. Copilot's paraphrased output included some sequences of texts copied verbatim from the article with only cosmetic alterations to others in a manner that retained the Chicago Tribune's original expression. Copilot's output to the user is reproduced below. The red text represents the text from Copilot's output that overlaps	1090. **Disputed - Not Supported by Cited Evidence; Not Supported by Admissible Evidence; Improper Legal Argument.** Microsoft disputes the fact contained in Paragraph 1090 and incorporates its response to Paragraph 1089. Microsoft further responds that News Plaintiffs' statement that Copilot's alleged "copying" was in a "manner that retained [Plaintiffs'] ... original expression"

with the original article. Ex. 806 [MSFT_AI_MDL_002882532] at -780-783. 	is improper legal argument and is not supported by the evidence. This fact states impermissible legal conclusions, including that Copilot "copied" text, which presumes the volitional conduct requirement for direct infringement, *see Cartoon*, 536 F.3d at 131, and that the output retained Plaintiffs' "original expression," which presumes both that the overlapping material is protected expression and that the output is substantially similar to it even after stripping out unprotectable elements like facts, *see Nihon Keizai Shimbun, Inc. v. Comline Business Data*, 166 F.3d 65, 70 (2d Cir. 1999). OpenAI defers to Microsoft's response.
1091. In another conversation (id: 51D[BingProd]8C2B55C89C373856685AF7673A08B0ACE8DEC4538274A57684FEEFBD9DC3F769) Copilot retrieved over 500 words from the Times article "Near the Zoo's snakes, a Hospital's Collection of Medicine" and proceeded to provide the user with a paraphrased version of the article. The text retrieved by Copilot is reproduced below. The red text represents the text from the RAG data retrieved by Copilot that overlaps with the original article. Ex. 806 [MSFT_AI_MDL_002882532] at -2889899-902.	1091. **Disputed - Improper Legal Argument; Not Supported by Cited Evidence; Not Supported by Admissible Evidence.** Microsoft disputes the fact contained in Paragraph 1091 and incorporates its response to Paragraph 1089 and 1090. OpenAI defers to Microsoft's response.

	
1092. Copilot's paraphrased output included some sequences of texts copied verbatim from the article with only cosmetic alterations to others in a manner that retained The Times's original expression. Copilot's output to the user is reproduced below. The red text represents the text from Copilot's output that overlaps with the original article. Ex. 806 [MSFT_AI_MDL_002882532] at -2889899-902.	1092. **Disputed - Improper Legal Argument; Not Supported by Cited Evidence; Not Supported by Admissible Evidence.** Microsoft disputes the fact contained in Paragraph 1092 and incorporates its response to Paragraph 1089 and 1090. OpenAI defers to Microsoft's response.

F. Defendants' Other Uses of News Content: New York Times Annotated Corpus, Anti-Regurgitation, Project Giraffe

1. OpenAI's Use of the New York Times Annotated Corpus to Develop a Commercial Product, in Violation of the User License Agreement

i) Restrictions on Using the New York Times Annotated Corpus under the User License Agreement

1093. The New York Times Annotated Corpus ("Annotated Corpus") is a compilation of over 1.8 million New York Times articles—specifically "nearly every article published in Times between January 01, 1987 and June 19, 2007." Ex. 808 [NYT_00146766] at -69.	1093. **Undisputed.** Microsoft defers to OpenAI.
1094. The Annotated Corpus includes not only the content of these Times articles but also "metadata provided by The New York Times Newsroom, The New York Times Indexing Service and the online production staff at NYTimes.com," such as information about each article's author. *Id.*	1094. **Undisputed.** Microsoft defers to OpenAI.

1095. The Annotated Corpus was previously made available by the Linguistic Data Consortium (LDC) at the University of Pennsylvania. Ex. 107 [NYT_00146736] at -36-37.	1095. **Undisputed.** Microsoft defers to OpenAI.
1096. [redacted] Ex. 107 [NYT_00146736] at -36.	1096. **Disputed in Part– Immaterial.** OpenAI responds as follows: The facts stated in Paragraph 1096 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Undisputed as to the language of the 2008 distribution agreement, which speaks for itself. News Ex. 107 at NYT_00146736. Pursuant to Fed. R. Evid. 106, according to the [redacted] corporate designee, [redacted] Ex. 191 ([redacted] 30(b)(6) Tr.) 56:15–57:7, 57:23–58:11. Microsoft defers to OpenAI.
1097. [redacted] Ex. 107 [NYT_00146736] at -44-45.	1097. **Disputed in Part– Immaterial.** OpenAI responds as follows: Undisputed as to the language of the 2008 distribution agreement, which speaks for itself. News Ex. 107 at NYT_00146744–45. However, the facts stated in Paragraph 1097 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Pursuant to Fed. R. Evid. 106,

	according to the [redacted] corporate designee, [redacted] Ex. 191 ([redacted] 30(b)(6) Tr.) 56:15–57:7, 57:23–58:11. Microsoft defers to OpenAI.
1098. The 2008 distribution agreement between The Times and LDC attached the "User License Agreement" that any "Permitted Researcher" would be required to sign before obtaining access to the Corpus. Ex. 107 [NYT_00146736] at -44-45.	1098. **Disputed in Part– Immaterial.** OpenAI responds as follows: Undisputed as to the language of the 2008 distribution agreement, which speaks for itself. News Ex. 107 at NYT_00146744–45. However, the facts stated in Paragraph 1098 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Pursuant to Fed. R. Evid. 106, according to the [redacted] corporate designee, [redacted]. Ex. 191 ([redacted] 30(b)(6) Tr.) 56:15–57:7, 57:23–58:11. Microsoft defers to OpenAI.

1099. This "User License Agreement" stated that the Annotated Corpus "may only be used for non-commercial linguistic education, research and technology development." Ex. 107 [NYT_00146736] at -44.	1099. **Disputed in Part– Immaterial.** OpenAI responds as follows: Undisputed as to the language of the user agreement in question, which speaks for itself. News Ex. 107 at NYT_00146744. However, the facts stated in Paragraph 1099 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Microsoft defers to OpenAI.
1100.The Times enforced the non-commercial use restriction for the New York Times Annotated Corpus. Ex. 1152 [Sandhaus_00000095] at -95; Ex. 309 [Procas 30(b)(6) Depo. Tr.] 269:21-270:5.	1100.**Disputed - Immaterial; Not Supported by Cited Evidence.** OpenAI responds as follows: The facts stated in Paragraph 1100 are immaterial and therefore disputed because these facts do not affect the outcome of the suit under the governing law. Further, pursuant to Fed. R. Evid. 106, according to the [redacted] corporate designee, [redacted] Ex. 191 ([redacted] 30(b)(6) Tr.) 77:16–22. [redacted]. *Id.* 55:11–14. [redacted] *Id.* 74:6–9. In fact, [redacted]. *Id.* 114:4–115:2. [redacted]

	[redacted]. *Id.* 92:10–15. Further, the facts set forth in Paragraph 1100 are disputed because they are not supported by the cited evidence. Specifically, Ms. [redacted] testimony is strong and precisely on point: [redacted] Ex. 191 ([redacted] 30(b)(6) Tr.) 55:11-14, 73:8-10, 74:6-9. And the Simone Procas testimony that [redacted] *See* Ex. 259 (S.Procas 30(b)(6) Tr.) 269:21–270:5; *see* also News Ex. 1152 at SANDHAUS_00000095. Microsoft defers to OpenAI.
1101. In January 2009 then-Times employee Evan Sandhaus was asked whether the Annotated Corpus could be "used to train possible commercial use machine learning algorithms"; he replied: "That kind of work you describe is not permitted under the LDC License. That being said, we are very open to discussing a commercial licensing arrangement to our data. Please follow up with me at [redacted] to discuss specifics." Ex. 1152 [Sandhaus_00000095] at -95.	1101. **Disputed– Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1101 are disputed because they are not supported by the cited evidence. Specifically, disputed that this email constitutes evidence of "enforcement" of non-commercial use restrictions for the Annotated Corpus. Undisputed as to the contents of the email from Mr. Sandhaus, which speaks for itself. Ex. 1152 (Sandhaus_00000095) at -95. OpenAI further disputes

	this fact to the extent it omits necessary context. Pursuant to Fed. R. Evid. 106, in an email [redacted] Ex. 10 at -042. Microsoft defers to OpenAI.

ii) OpenAI's Acquisition of the New York Times Annotated Corpus

1102. OpenAI acquired a copy of the Annotated Corpus on October 24, 2019, when OpenAI employee Long Ouyang purchased a copy from LDC and signed the User License Agreement on behalf of OpenAI. Ex. 1153 [LDC-NYT_000033] at -33-34.	1102. **Disputed.** OpenAI responds as follows: The facts set forth in Paragraph 1102 are disputed as misleading for omitting necessary context. *See* News Ex. 1153 at LDC-NYT_000033–34. Pursuant to Fed. R. Evid. 106, to the best of Mr. Ouyang's knowledge, in light of the commercial use prohibition in the User Agreement for The New York Times's Annotated Corpus, OpenAI never used the New York Times Annotated Corpus. Ex. 248 (L. Ouyang 30(b)(1) Tr.) 23:10–16, 133:18–24. Mr. Ouyang's own testimony confirms he initially intended to use the Corpus to train a model but did not use it to his knowledge. *Id.* Microsoft defers to OpenAI.

1103. Mr. Ouyang paid the fee for the Corpus with OpenAI funds. Ex. 891 [OPCO_MDL_001899968] at -68; Ex. 661 [Ouyang Depo. Tr.] at 135:7-19.	1103. **Disputed.** OpenAI responds as follows: The facts set forth in Paragraph 1103 are disputed as misleading for omitting necessary context. *See* News Ex. 891 at OPCO_MDL_001899968; Ex. 248 (L. Ouyang 30(b)(1) Tr.) 135:7–19. Pursuant to Fed. R. Evid. 106, to the best of Mr. Ouyang's knowledge, in light of the commercial use prohibition in the User Agreement for The New York Times's Annotated Corpus, OpenAI never used the New York Times Annotated Corpus. Ex. 248 (L. Ouyang 30(b)(1) Tr.) 23:10–16, 133:18–24. Microsoft defers to OpenAI.
1104. Prior to purchasing a copy of the Annotated Corpus, Mr. Ouyang sought and obtained approval from OpenAI to use his company-issued credit card to pay the $300 fee. Ex. 891 [OPCO_MDL_001899968] at -68; Ex. 661 [Ouyang Depo. Tr.] at 135:7-19.	1104. **Disputed.** OpenAI responds as follows: The facts set forth in Paragraph 1104 are disputed as misleading for omitting necessary context. *See* Ex. 891 (OPCO_MDL_001899968) at -68; Ex. 661 (L. Ouyang 30(b)(1) Tr.) 135:7–19. Pursuant to Fed. R. Evid. 106, to the best of Mr. Ouyang's knowledge, in light of the commercial use prohibition in the User Agreement for The New York Times's Annotated Corpus, OpenAI never used the New York Times Annotated Corpus. Ex. 248 (L. Ouyang 30(b)(1) Tr.) 23:10–16, 133:18–24. Disputed as misleading for omitting necessary context. Microsoft defers to OpenAI.
1105. By signing the User License Agreement for the Annotated Corpus, OpenAI agreed that it would not use the Corpus for commercial purposes. The Agreement provided	1105. **Disputed in Part.** OpenAI responds as follows: Undisputed as to the language of the User License Agreement, which speaks for itself. News Ex. 1153 at LDC-

that the "Data may only be used for non-commercial linguistic education, research and technology development" and that "User shall not publish, retransmit, display, redistribute, reproduce, or commercially exploit the Data in any form." Ex. 1153 [LDC-NYT_000033] at -33.	NYT_000033. However, disputed as misleading for omitting necessary context. Pursuant to Fed. R. Evid. 106, to the best of Mr. Ouyang's knowledge, in light of the commercial use prohibition in the User Agreement for The New York Times's Annotated Corpus, OpenAI never used the New York Times Annotated Corpus. Ex. 248 (L. Ouyang 30(b)(1) Tr.) 23:10–16, 133:18–24. Microsoft defers to OpenAI.
1106. In an email exchange about OpenAI's purchase of a copy of the Annotated Corpus, an LDC employee reminded Mr. Ouyang that "the Special License Agreement for this corpus prohibits commercial development under all circumstances." Ex. 1260 [LDC-NYT_000046] at -46.	1106. **Disputed.** OpenAI responds as follows: Undisputed as to the language of the email, which speaks for itself. News Ex. 1260 at LDC-NYT_000046. However, disputed as misleading for omitting necessary context. Pursuant to Fed. R. Evid. 106, to the best of Mr. Ouyang's knowledge, in light of the commercial use prohibition in the User Agreement for The New York Times's Annotated Corpus, OpenAI never used the New York Times Annotated Corpus. Ex. 248 (L. Ouyang 30(b)(1) Tr.) 23:10–16, 133:18–24. Microsoft defers to OpenAI.
1107. Mr. Ouyang, on behalf of OpenAI, reviewed the license before making the purchase and downloading it" and he determined that it "would not be appropriate" to use the Annotated Corpus "to train a model." Ex. 661 [Ouyang Depo. Tr.] at 23:10-13, 24:1-9.	1107. **Undisputed.** Microsoft defers to OpenAI.

1108. Under the User License Agreement, OpenAI did not have permission to use the Annotated Corpus for commercial purposes, and OpenAI would be violating the Agreement if it used the Annotated Corpus "in the development of a commercial large language model" or for "commercial technology." Ex. 108 [DiPersio Depo. Tr.] at 156:19-23, 164:24-165:7, 168:8-21. *See also id.* at 167:25-168:6 (confirming that the User License Agreement was "the only agreement that applied to OpenAI's use of any New York Times content from the LDC."); Ex. 1153 [LDC-NYT_000033] at -33.	1108. **Disputed in Part - Improper Legal Argument.** OpenAI responds as follows: Undisputed as to the language of the User License Agreement, which speaks for itself. News Ex. 1153 at LDC-NYT_000033. However, disputed to the extent this fact asserts an improper legal conclusion regarding the agreement. Further, the Court should disregard the disputed portion of Paragraph 1108 because the assertion constitutes improper attorney argument and/or is conclusory. It thus requires no response. Microsoft defers to OpenAI.
1109. Mr. Ouyang immediately brought the New York Times Annotated Corpus to OpenAI engineers developing OpenAI's commercial GPT-3 large language model. Ex. 1253 [OPCO_MDL_000495790] at -90.	1109. **Disputed in Part.** OpenAI responds as follows: Disputed as to the characterization of Mr. Ouyang's message as "immedia[te]" and misleading for omitting necessary context. *See* News Ex. 1253 at OPCO_MDL_000495790. Pursuant to Fed. R. Evid. 106, to the best of Mr. Ouyang's knowledge, in light of the commercial use prohibition in the User Agreement for The New York Times's Annotated Corpus, OpenAI never used the New York Times Annotated Corpus. Ex. 248 (L. Ouyang 30(b)(1) Tr.) 23:10–16, 133:18–24. Microsoft defers to OpenAI.
1110. On October 25, 2019, one day after purchasing the Corpus and signing the User License Agreement, Mr. Ouyang informed OpenAI colleague Ben Mann that he had	1110. **Disputed in Part.** OpenAI responds as follows: Undisputed as to the language of the Slack message, which speaks for itself. News Ex. 1253 at OPCO_MDL_000495790. However, disputed as

"purchased" the "NYT Annotated Corpus." Ex. 1253 [OPCO_MDL_000495790] at -90.	misleading for omitting necessary context. Pursuant to Fed. R. Evid. 106, to the best of Mr. Ouyang's knowledge, in light of the commercial use prohibition in the User Agreement for The New York Times's Annotated Corpus, OpenAI never used the New York Times Annotated Corpus. Ex. 248 (L. Ouyang 30(b)(1) Tr.) 23:10–16, 133:18–24. Microsoft defers to OpenAI.

iii) The OpenAI "Nest" Team and their Use of the New York Times Annotated Corpus to Develop the GPT-3 Large Language Model

1111. In October 2019, the "Nest" project was an OpenAI codename for the GPT-3 model. Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 18:13-17.	1111. **Undisputed.** Microsoft defers to OpenAI.
1112. The "Nest" project entailed "developing the technology that [went] into building GPT-3," and this project "eventually resulted in a couple of different models being trained under the GPT-3 model family, moniker." Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 19:21-24, 21:1-6.	1112. **Undisputed.** Microsoft defers to OpenAI.
1113. The Nest team trained OpenAI's GPT-3 model. Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 76:4-20; Ex. 625 [Amodei Depo. Tr.] at 259:2-18 ("the original effort to train that model [GPT-3] was known as the Nest project," and the Nest Team "was the primary team responsible for training the GPT-3 model.").	1113. **Undisputed.** Microsoft defers to OpenAI.

1114. The "Nest" team understood their work would be used for "commercial applications." Ex. 893 [OPCO_MDL_001006273] at –74.	1114. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1114 are disputed because they are unsupported by the record. News Ex. 893 is a document titled "Dario's Role at OpenAI," which, by its own terms, "attempts to capture a conversation between Dario and Sam about [Dario's] role at OpenAI" and sets forth "a proposed agreement about Dario's role going forward." *Id.* OPCO_MDL_001006276. The document states only that "[w]orking on Nest includes collaborating with Mira on commercial applications of Nest for Microsoft" as part of Dario's own role (*id.* OPCO_MDL_001006277) and contains no statement reflecting the understanding of the broader Nest engineering team. Microsoft defers to OpenAI.
1115. The "critical work" the "nest team" was doing to support "our commercialization work with Microsoft" included "unlock[ing] the second $1B investment" from Microsoft. Ex. 895 [OPCO_MDL_004625973] at -73.	1115. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1115 are disputed because they are not supported by the cited evidence. Specifically, disputed as to the characterization of the email from Mira Murati dated January 29, 2020, which speaks for itself and does not use the word "support." Rather, Murati's email states "It's important to call out that we are building this critical work on top of the unique and valuable work done by the language team, nest team, and infra team." *See* News Ex. 895 at OPCO_MDL_004625973.

	Microsoft defers to OpenAI.
1116. The "Nest" team used the New York Times Annotated Corpus to develop OpenAI's commercial GPT-3 large language model. Ex. 897 [OPCO_MDL_000479314] at -14; Ex. 899 [OPCO_MDL_001686982] at -83; Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 21:20-22:14 (On October 27, 2019, three days after Mr. Ouyang purchased the Annotated Corpus for OpenAI, Mr. Mann commented on a "Nest Presentation" document discussing the "200B parameter model" and plans to "Fine-tune[]" the model "to New York Times," noting that Mr. Ouyang had just purchased the Annotated Corpus.). *See also* Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 24:13-24 (The reference to the "200B parameter model" was referring to "GPT-3," which was "a 170-something billion parameter model, but while [OpenAI was] working on it, [they] always shorthanded that to 200 billion.").	1116. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1116 are disputed because they are not supported by the cited evidence. News Ex. 897 is a Google Docs comment thread on a document titled "Deliverables for Possible Nest Presentation 11/22," in which, under a comment heading "Fine-tuned to New York Times," Ben Mann wrote "long just purchased this," and Dario Amodei replied "Awesome!" This reflects, at most, a brainstorming discussion about a possible future use of the Annotated Corpus, not a statement that the Nest team in fact used the Annotated Corpus to develop GPT-3. Mr. Mann testified that, [redacted] Ex. 236 (B. Mann 30(b)(1) Tr. (Feb. 20, 2026)) 51:18-52:2. Plaintiffs' cited evidence thus does not establish that the Annotated Corpus was in fact used to develop GPT-3, as opposed to merely being discussed, prepared, or proposed as a possible fine-tuning dataset. Microsoft defers to OpenAI.
1117. OpenAI used the New York Times Annotated Corpus for the fine-tuning phase of training the GPT-3 model. Ex. 899 [OPCO_MDL_001686982] at -83; Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 36:8-13 (Q: "Fair to say	1117. **Disputed — Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that News Exs. 899, 901, 903, 625, and 663 contain the referenced internal messaging exchanges and deposition

this document is discussing the possibility of using The New York Times Annotated Corpus to train these models?" A: "It seems like it's discussing that, yes."); Ex. 901 [OPCO_MDL_000494603] at -03 (Mr. Amodei in November 2019 explaining his plans to "fine-tune" on the "NYT dataset" which Long Ouyang had "purchas[ed]"); Ex. 625 [Amodei Depo. Tr.] at 302:2-10 (Amodei testifying that this document "would seem to suggest that I did" fine-tune the model using the Annotated Corpus); Ex. 903 [OPCO_MDL_002139639] at -41 (Mr. Amodei in a March 2020 discussion within the "nest" team Slack channel referencing "our generative fine-tuning" to "NYT"); Ex. 625 [Amodei Depo. Tr.] at 313:6-11 (Q: "you did do fine-tuning using The New York Times dataset?" A: "that would appear to be suggested by the text, yes.").	testimony. However, disputed that the cited evidence supports Plaintiffs' characterization that OpenAI used the New York Times Annotated Corpus for the fine-tuning phase of training the GPT-3 model, and disputed the inferences Plaintiffs draw from that evidence as contradicted by the record. News Exs. 899 and 901 discuss early, preliminary possibilities and planning proposals written well prior to the training of GPT-3, and do not document actual base model training. Ex. 236 (B. Mann 30(b)(1) Tr. (Feb. 20, 2026)) 36:8–13. News Ex. 903 refers generally to preliminary "generative fine-tuning" experiment results within the early "nest" model family, but does not state or indicate that any such results involved the GPT-3 model. GPT-3 did not undergo any fine-tuning; a core technological breakthrough of the GPT-3 model was its ability to perform task evaluation zero-shot or few-shot without fine-tuning. News Ex. 486 (GPT-3 Paper) at 5 ("We emphasize that these 'learning' curves involve no gradient updates or fine-tuning, just increasing numbers of demonstrations given as conditioning."). Both Mr. Mann and Dr. Amodei testified that [redacted] Ex. 164 (D. Amodei 30(b)(1) Tr.) 298:24-299:4, 302:2-10, 316:5-317:17 ([redacted]); Ex. 236 (B. Mann 30(b)(1) Tr.

	(Feb. 20, 2026)) 24:16-27:6 [redacted], 51:18-52:6. Microsoft defers to OpenAI.
1118. The purpose of fine-tuning the GPT-3 model on the New York Times Annotated Corpus was to improve the model's performance on news content. Ex. 663 [Mann Feb. 20, 2026 Depo Tr.] at 31:22-33:4 ("in the fine-tuning step, you specialize a model to a specific task" and so "for fine-tuning to something like The New York Times,"; "the model is trying to predict the next token in text that would have come from New York Times articles."); *see also* Ex. 665 [Pachocki Depo. Tr.] at 256:4-9 ("Fine-tuning is the process of taking an existing model that has presumably already been pretrained on some data and training it on a smaller dataset to cause it to behave in some way – or to make it better at predicting specific kind of data").	1118. **Disputed.** OpenAI responds as follows: This fact is disputed because the premise of Paragraph 1118, that GPT-3 was fine-tuned on the New York Times Annotated Corpus, is not supported by the record. As OpenAI explained in the paper introducing GPT-3, "Language Models are Few-Shot Learners," GPT-3 was "applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model." News Ex. 486 (GPT-3 Paper) at Abstract. Benjamin Mann testified that, [redacted] Ex. 236 (B. Mann 30(b)(1) Tr.) 32:14-33:4, 51:18-52:2. Because the cited evidence does not establish that GPT-3 was fine-tuned on the Annotated Corpus at all, it cannot establish any purpose

	for doing so, including the purpose alleged in Paragraph 1118. *See* Response to Fact 1117. Microsoft defers to OpenAI.
1119. OpenAI also used the New York Times Annotated Corpus to help filter Common Crawl data for use in training OpenAI's models. Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 125:11-17 ("I think it is fair to say that OpenAI used some documents from WebText and The New York Times to [redacted] filter Common Crawl for documents which [redacted] similar to those documents with the intent of finding legible documents and not spam and junk."); Ex. 743 [OPCO_MDL_000479924] at -25 (July 2021 document describing work on the "Nest" project, including how "Nick Ryder created a new Logistic Regression (LR) filter by reviving Ben Mann's old code" and noting this filter "was trained with a mix of webtext, wikipedia, nytimes, and libgen documents"); Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 117:12-120:23 (the "intent" of this filter was to find content in Common Crawl that was "similar" to Times content).	1119. **Disputed in Part.** OpenAI responds as follows: Undisputed that text from the New York Times Annotated Corpus was part of a mixture of documents used to train an out-of-scope classifier model to assist in filtering raw Common Crawl html data. *See* Ex. 267 (N. Ryder 30(b)(6) Tr. (Jan. 29, 2026)) 125:11–17; News Ex. 743 at OPCO_MDL_000479925. However, disputed as to the characterization that all of OpenAI's models were trained using the New York Times Annotated Corpus. The disputed proposition is not supported by the cited evidence. Specifically, pursuant to Fed. R. Evid. 106, OpenAI's corporate designee testified that this filter, known as the logistic regression filter (the "LR filter"), was designed to distinguish human-authored entries in Common Crawl from entries that were suspected to be spam or autogenerated. *See* Ex. 250 (A. Paino 30(b)(6) Tr.) 243:7–246:21. Building a classifier to filter Common Crawl data for pretraining is part of the pretraining process—it is a necessary step in assembling the training dataset. The resulting Common Crawl dataset was used in an iteration on GPT-3; [redacted] *Id.* 246:22–248:6. Microsoft defers to OpenAI.

1120. "OpenAI used the [NYT Annotated] Corpus to filter and select which content from Common Crawl to include as training data for the model." Goldstein Opening Rep. ¶ 249.	1120. **Disputed in Part.** OpenAI responds as follows: Undisputed that text from the New York Times Annotated Corpus was part of a mixture of documents used to train an out-of-scope classifier model to assist in filtering raw Common Crawl html data. However, disputed that Plaintiffs' characterization of this process as OpenAI having "used" the Corpus "to filter and select which content from Common Crawl to include as training data for the model" accurately reflects the record. OpenAI's corporate designee testified that this filter, known as the logistic regression filter (the "LR filter"), was designed to distinguish human-authored entries in Common Crawl from entries that were suspected to be spam or autogenerated. Ex. 250 (A. Paino 30(b)(6) Tr.) 243:7-246:21. The LR filter was not designed to "select" data from Common Crawl that appeared similar to Times articles; rather, it was designed to identify whether Common Crawl documents were likely human-authored rather than unreadable data, repetitive text, or machine-generated spam. Ex. 250 (A. Paino 30(b)(6) Tr.) 246:22-248:6. Microsoft defers to OpenAI.
1121. OpenAI also used the Annotated Corpus as a "test set" to evaluate the performance of the GPT-3 model. Ex. 905 [OPCO_MDL_000479741] at -41 (March 2020 document listing the "NYT test set" as one of the "generative test sets" OpenAI was using for the Nest project); Ex. 743 [OPCO_MDL_000479924] at -27-28 (listing "new-york-	1121. **Disputed in Part.** Undisputed that OpenAI used The New York Times Annotated Corpus. Disputed that it was used to evaluate the performance of the GPT-3 model. OpenAI's corporate designee, Nicholas Ryder, testified that "OpenAI included The New York Times's Annotated Corpus in an eval suite" with other evals and the

times_loss" as an "eval[] used" for the Nest project); Ex. 621 [Ryder 30(b)(6) Depo. Tr.] at 121:5-12 ("OpenAI included The New York Times's Annotated Corpus in an eval suite. . . . And aggregate evals were kind of -- were looked at to evaluate different ablations and design decisions."); Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 67:12-17 ("we were using the test sets to judge whether our model seems to be making progress in its generalization abilities").	aggregate evals "were looked at to evaluate different ablations and design decisions." Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr. (Jan. 29, 2026)) 121:5-12. When asked whether "OpenAI was using The New York Times's Annotated Corpus to evaluate whether New York Times data would have improved performance of the model," Ryder responded "[n]o." Hurst Decl. Ex. 104 (N. Ryder 30(b)(6) Tr. (Jan. 29, 2026)) 122:8-13. Microsoft defers to OpenAI.
1122. "Test sets" were "helpful" in "building the Nest models" and "useful" in helping OpenAI "complete the GPT-3 model." Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 68:14-22.	1122. **Disputed in Part.** OpenAI responds as follows: Undisputed that the cited document contains testimony that "test sets" were useful in building the nest models. However, disputed to the extent that News Plaintiffs' attempts to characterize this testimony as referring to any particular "test set." Pursuant to Fed. R. Evid. 106, Mr. Mann also testified [redacted] Ex. 236 (B. Mann 30(b)(1) Tr. (Feb. 20, 2026)) 67:5-68:3. Microsoft defers to OpenAI.
1123. OpenAI used the Annotated Corpus to assess the performance of the GPT-3 model and to train and develop the GPT-3 model. Goldstein Opening Rep. ¶¶ 244-46; *see also* Fact 1117 (OpenAI used the New York Times Annotated	1123. **Disputed.** OpenAI responds as follows: This fact is disputed because OpenAI did not use the Annotated Corpus to "train and develop" the GPT-3 model. Both Mr. Mann and Dr. Amodei testified [redacted]

Corpus for the fine-tuning phase of training the GPT-3 model. Ex. 899 [OPCO_MDL_001686982] at -83; Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 36:8-13 (Q: "Fair to say this document is discussing the possibility of using The New York Times Annotated Corpus to train these models?" A: "It seems like it's discussing that, yes."); Ex. 901 [OPCO_MDL_000494603] at -03 (Mr. Amodei in November 2019 explaining his plans to "fine-tune" on the "NYT dataset" which "Long Ouyang" had "purchas[ed]"); Ex. 625 [Amodei Depo. Tr.] at 302:2-10 (Amodei testifying that this document "would seem to suggest that I did" fine-tune the model using the Annotated Corpus); Ex. 903 [OPCO_MDL_002139639] at -41 (Mr. Amodei in a March 2020 discussion within the "nest" team Slack channel referencing "our generative fine-tuning" to "NYT"); Ex. 625 [Amodei Depo. Tr.] at 313:6-11 (Q: "you did do fine-tuning using The New York Times dataset?" A: "that would appear to be suggested by the text, yes.").)	[redacted] Ex. 164 (D. Amodei 30(b)(1) Tr.) 298:24-299:4, 302:2-10; Ex. 236 (B. Mann 30(b)(1) Tr. (Feb. 20, 2026)) 51:18-52:6. Microsoft defers to OpenAI.

iv) OpenAI's Commercial API Product and its Reliance on GPT-3

1124. The initial version of the OpenAI API was released in June 2020. Ex. 505 [Lightcap 30(b)(6) Depo Tr.] at 15:5-10; Ex. 907 [OPCO_MDL_001692573] at -75 (June 11, 2020 public announcement from OpenAI: "Today the API runs models with weights from the GPT-3 family").	1124. **Undisputed.** Microsoft defers to OpenAI.
1125. The initial version of the OpenAI API was powered by the GPT-3 model. Ex. 505 [Lightcap 30(b)(6)	1125. **Undisputed.**

Depo. Tr.] at 15:5-10; Ex. 907 [OPCO_MDL_001692573] at -75 (June 11, 2020 public announcement from OpenAI: "Today the API runs models with weights from the GPT-3 family").	Microsoft defers to OpenAI.
1126. The release of the OpenAI API on June 11, 2020 allowed developers to access the GPT-3 model. Ex. 505 [Lightcap 30(b)(6) Depo. Tr.] at 15:8-10; Ex. 907 [OPCO_MDL_001692573] at -77 ("We're releasing an API for accessing new AI models developed by OpenAI").	1126. **Undisputed.** Microsoft defers to OpenAI.
1127. When OpenAI released the OpenAI API on June 11, 2020, OpenAI described the API as a "commercial product." Ex. 907 [OPCO_MDL_001692573] at -77.	1127. **Undisputed.** However, OpenAI responds to further note that the cited document explains that the OpenAI API was launched on June 11, 2020 as a restricted "private beta rather than general availability" to allow OpenAI to research safety aspects, build content controls, and work closely with approved partners before broader commercial deployment *See* News Ex. 907 at OPCO_MDL_001692574–76. Microsoft defers to OpenAI.
1128. OpenAI earns revenue from the OpenAI API. Ex. 505 [Lightcap 30(b)(6) Depo. Tr.] at 14:6-10 ("We charge per token. So you can call our API, and the unit of account is a token, which is kind of like the unit of account for a language model. And then depending on the number of tokens consumed, we pass the bill to the customer, and they pay us."); *id.* at 15:21-23 ("[W]e have many millions of accounts that have signed up, developer accounts signed up	1128. **Undisputed.** Microsoft defers to OpenAI.

for our API platform."); Ex. 501 [Lightcap 30(b)(6) Depo. Ex. 1] (financial statement showing OpenAI's revenues from the API from 2020 through 2024).	

v) OpenAI's Removal of Copyright Management Information from the New York Times Annotated Corpus

1129. OpenAI produced a file in this case labeled OPCO_MDL_004389835, which contains 1,855,658 New York Times articles. Goldstein Opening Rep. ¶ 250; Ex. 909 [OPCO_MDL_004389835].	1129. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OPCO_MDL_004389835 contains 1,855,658 records. However, disputed that those records are New York Times articles. For that reason, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
1130. The Annotated Corpus contains 1,855,658 Times articles. Ex. 808 [NYT_00146766] at -86.	1130. **Disputed in Part - Not Supported by Cited Evidence.** Defendants respond as follows: Undisputed that the cited document indicates that The New York Times Annotated Corpus claimed that it contained 1,855,658 documents. However, disputed to the extent that The New York Times characterizes that statement as providing any indication as to whether the content of those documents are owned or copyrighted to The Times. The disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
1131. [Intentionally left blank].	1131. [Intentionally left blank].

1132. The file produced by OpenAI labeled OPCO_MDL_004389835 is a json file. Goldstein Opening Rep. ¶ 251; Ex. 909 [OPCO_MDL_004389835].	1132. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1132 are disputed because the cited evidence is incorrect. Specifically, the file produced by OpenAI labeled OPCO_MDL_004389835 is a ZIP file, not a "json file." Microsoft defers to OpenAI.
1133. After Mr. Ouyang purchased the New York Times Annotated Corpus in October 2019, Mr. Mann recommended that they "use json" to process the Corpus. Ex. 1253 [OPCO_MDL_000495790] at -90.	1133. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that the cited document includes the language "use json." Disputed that the document reflects that json was used "to process the Corpus." Therefore, the disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
1134. A former OpenAI employee named Ben Mann "may have created the 'new-york-times' directory" where the OPCO_MDL_004389835 file was stored and found. Ex. 902 [April 8, 2026 Declaration of OpenAI Associate General Counsel Michael Trinh] ¶ 3.	1134. **Undisputed.** Microsoft defers to OpenAI.
1135. OpenAI produced its copy of the New York Times Annotated Corpus in this case, and that file is labeled OPCO_MDL_004389835. Goldstein Opening Rep. ¶ 250; Ex. 909 [OPCO_MDL_004389835].	1135. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed that the file labeled OPCO_MDL_004389835 contains a "copy" of the New York Times Annotated Corpus. Therefore, the proposition set forth in Paragraph 1135 is not supported by the cited evidence.

	Microsoft defers to OpenAI.
1136. OpenAI removed copyright management information from the New York Times Annotated Corpus. Goldstein Opening Rep. ¶ 255; Ex. 909 [OPCO_MDL_004389835]; Ex. 901 [OPCO_MDL_000494603] at -03; Ex. 663 [Mann Feb. 20, 2026 Depo. Tr.] at 49:1-3.	1136. **Disputed.** OpenAI responds as follows: OpenAI's text extractors do not "remove" anything. Rather, they extract material that is suitable for model training. Not extracting data is not equivalent to removing data. *See, e.g.*, ECF 1748 (C. Callison-Burch Decl.) ¶ 88 ("Importantly, the purpose of the text extraction process used by OpenAI and other NLP practitioners is to identify text to keep, not text to remove. The scripts are designed to extract readable natural language text from HTML files for use as training data. Any loss of CMI during this process is an incidental byproduct of the text extraction methodology—it is not a purposeful or intentional removal of CMI."); Fed. R. Evid. 106. Microsoft defers to OpenAI.
1137. The file produced by OpenAI labeled OPCO_MDL_004389835 contains the text of New York Times articles but is missing metadata that initially accompanied the Corpus. Goldstein Opening Rep. ¶ 258; Ex. 808 [NYT_00146766] at -69-70.	1137. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: Plaintiffs' assertion is not supported by the cited evidence. Dr. Goldstein did not identify any New York Times articles in the cited evidence. Further disputed as misleading: Dr. Goldstein's report stated that it did not include "all of the metadata," conceding that some metadata was still present. ECF 1819-1 (T. Goldstein Decl. Ex. A) ¶ 258; Microsoft defers to OpenAI.

1138. When OpenAI processed the New York Times Annotated Corpus in November 2019, it "didn't include the summaries, annotations, or most of the metadata." Ex. 901 [OPCO_MDL_000494603] at -03; Ex. 663 [Mann Feb. 20, 2026 Depo Tr.] at 49:1-3 (addressing Ex. 901 [OPCO_MDL_000494603] at -03: "I would say it wasn't a removal, per se, and more of a just not including it").	1138. **Undisputed.** Microsoft defers to OpenAI.

2. OpenAI's Use of News Content for Project Giraffe

1139. OpenAI deploys an inference time Bloom filter called the Giraffe filter.	1139. **Undisputed.** Microsoft defers to OpenAI.
1140. The Giraffe filter was intended to detect and block regurgitation of certain copyrighted material identified on certain blocklists. Ex. 911 [OPCO_MDL_004820692] at -92 ("To prevent copyright regurgitation, we have deployed mitigations at various stages. SafeSys mainly worked on two mitigations under the code name 'Giraffe': inference-time bloom filter (BF) blocking and post-training intervention."); Ex. 740 [OPCO_MDL_004809431] at -36 ("At the inference time, we deployed a distributed, large-scale bloom filter to remove n-gram phrases that match any known copyrighted materials in our database."); Ex. 913 [OPCO_MDL_004661064] at 23.	1140. **Undisputed.** Microsoft defers to OpenAI.

1141. OpenAI launched ChatGPT without a regurgitation filter. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 263:16-23.	1141. **Undisputed.** Microsoft defers to OpenAI.
1142. OpenAI discussed how "heavy handed interventions" to prevent regurgitation produced "a very unusable product" and worried that training the model to never output training data would cause "big performance regressions on factuality"‘ Ex. 915 [OPCO_MDL_004411433] at -53; Ex. 917 [OPCO_MDL_001699915] at -18.	1142. **Disputed in Part.** OpenAI responds as follows: Undisputed that the quoted text appears in the cited evidence. However, disputed to the extent News Plaintiffs' assertion implies these statements concern general interventions to prevent training data regurgitation. News Ex. 915 shows employees were discussing "[OpenAI's] first heavy handed interventions handicapping the model's quote ability" which were quoting limitations for specific news publishers, such as The New York Times. News Ex. 915 at OPCO_MDL_004411453. New Plaintiffs are taking the quotes out of context. For example, the comment regarding "heavy handed interventions" resulting in "a very unusable product" (*Id.*) referred specifically to domain-blocking and quoting limitations for specific news publishers (such as The New York Times), not general interventions to prevent training data regurgitation. Further disputed as statements reflecting the personal knowledge and opinions of OpenAI employees are insufficient to establish OpenAI's discussions about regurgitation and outputs generally. Microsoft defers to OpenAI.
1143. OpenAI did not deploy the initial version of Giraffe filter to ChatGPT until a few weeks after ChatGPT's	1143. **Undisputed.**

November 30, 2022 launch. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 212:6-16.	Microsoft defers to OpenAI.
1144. The first version of the Giraffe filter was an "in-memory" filter and included a limited number of copyrighted materials such as Harry Potter, the famous "Quake inverse square root" code snippet, and snippets from approximately 20 books. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 299:9-300:14.	1144. **Undisputed.** Microsoft defers to OpenAI.
1145. The version of the Giraffe filter put in place several weeks after the launch of ChatGPT did not include content related to The Times or the *Daily News* Plaintiffs. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 213:18-214:3.	1145. **Undisputed.** Microsoft defers to OpenAI.
1146. The Giraffe filter applied to both ChatGPT and OpenAI's API. Ex. 667 [Monaco Jan. 27, 2026 30(b)(6) Depo. Tr.] at 30:14-18.	1146. **Undisputed.** Microsoft defers to OpenAI.
1147. Immediately after the initiation of these lawsuits, OpenAI searched for and copied News Plaintiffs' content it believed most likely to be output by its LLMs to build a Bloom filter to block output of that content. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 220:2-16.	1147. **Disputed in Part - Not Supported by Cited Evidence; Improper Legal Argument.** OpenAI responds as follows: Undisputed that a dataset was prepared in late December 2023 containing training data content that was related to The New York Times. However, the rest of the paragraph is disputed because it is not supported by the cited evidence. Specifically, disputed that the dataset was created based on "content [OpenAI] believed most likely to be output by its LLMs." The cited testimony references the existence of the dataset but does not discuss the selection process for that dataset. Further

	disputed to the extent that the content was "copied" for purposes of copyright infringement. That assertion calls for a legal conclusion as to whether content was "copied," and the Court should therefore disregard it as improper attorney argument. For the same reason, disputed to the extent that "content it believed most likely to be output" is a legal conclusion regarding OpenAI's state of mind. Microsoft defers to OpenAI.
1148. OpenAI did not add n-grams from *New York Times* articles and *Daily News* Plaintiffs' until each Plaintiff filed their respective lawsuits. Ex. 651 [Monaco Apr., 8, 2026 30(b)(6) Depo. Tr.] at 235:25-236:9, 248:9-25.	1148. **Undisputed.** Microsoft defers to OpenAI.
1149. OpenAI has never added n-grams from Ziff Davis articles to the Giraffe filter. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 292:12-18; Delorey Rebuttal Rep. ¶ 117; Delorey Reply Rep. ¶ 44.	1149. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI's witness testified that he was unaware of an "effort to add content related to Ziff Davis specifically." Ex. 240 (V. Monaco 30(b)(6) Tr. (Apr. 8, 2026)) 292:17-18. However, it is disputed that the Giraffe filter does not contain any n-grams from Ziff Davis articles, as the assertion is not supported by the cited evidence. Microsoft responds that these facts do not pertain to claims asserted against Microsoft.
1150. OpenAI corporate representative who designated on topics related to the Giraffe filter testified as follows: "Q: Do you know if [Ziff Davis's] content has been	1150. **Undisputed.**

added to the Bloom filter? A: I'm not aware of an effort to add content related to Ziff Davis specifically."). Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 292:15-18.	Microsoft responds that these facts do not pertain to claims asserted against Microsoft.
1151. OpenAI did not create any entity-specific datasets for its Giraffe filter except for certain entities that had sued it. *See*. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 230:7-231:6.	1151. **Undisputed.** Microsoft defers to OpenAI.
1152. OpenAI could not use the Giraffe filter to block regurgitation of all copyrighted training data because adding all of the training data to the Giraffe filter "would have broken the product." Ex. 760 [Monaco Jan. 28, 2026 30(b)(6) Depo. Tr.] at 43:18-23.	1152. **Disputed.** The cited evidence does not support this statement, which misleadingly characterizes the quoted testimony. The record likewise does not support the inference that training data not added to the filter would otherwise be subject to regurgitation. The balance of evidence demonstrates that regurgitation of training data occurs rarely. OpenAI also disputes the implication that the contents of the Bloom filter were limited only by size constraints, the interpretation of "broken the product," and the suggestion that the stated reason was the only reason OpenAI did not add all training data to the Giraffe Bloom filter. The Giraffe Bloom filter uses a probabilistic data structure to compare strings of characters, known as n-grams, from model completions (i.e., outputs in response to a prompt) against a known dataset of n-grams. If the filter suggests a match, the model blocks further display of the output. Bloom filters have "perfect recall," so they catch every true match without false negatives, but they can produce false positives by blocking benign outputs (e.g.,

	public domain text such as Bible passages or famous quotes). As the filter's dataset grows, false positives increase and significant usability issues (such as delays in streaming) ensue. OpenAI therefore selected for the Giraffe Bloom filter the content most susceptible to potential regurgitation (despite other safeguards), while keeping the filter's size within operational limits. Plaintiffs fail to acknowledge that the Giraffe Bloom filter is considered a copyright mitigation layer of last resort and is one of many anti-regurgitation mechanisms OpenAI employs. *See* ECF 1748 (C. Callison-Burch Decl.) ¶ 111; News Ex. 483 (C. Callison-Burch Opening Rep.) ¶¶ 420-424; News Ex. 675 (C. Callison-Burch Reply Rep.) ¶¶ 93-95. Microsoft defers to OpenAI.
1153. OpenAI executives approved shipping a GPT-4 model with a known regurgitation problem rather than delaying its commercial launch, concluding that commercial timing justified shipping the model. Ex. 919 [OPCO_MDL_001655887] at -87.	1153. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1153 are disputed because the cited evidence is incorrect. Specifically disputed, for several reasons. First, Plaintiffs confuse the "52B model" (i.e., an internal code name for GPT-3.5) being discussed in the cited exhibit with GPT-4, which was not released until more than a year after the date of the document. A. Blanco Opp. Decl. Ex. 144 (OPCO_MDL_001876547); News Ex. 483 (C. Callison-Burch Opening Rep.) ¶¶ 254-255. Second, although the cited exhibit obliquely references a "regurgitation issue" that was in the process of being fixed,

	the document does not provide evidence that OpenAI approved shipping a model with a "known regurgitation problem." The cited exhibit also does not support the contention that "commercial timing justified shipping" this unspecified version of GPT-3.5 (not GPT-4). To the contrary and pursuant to Fed. R. Evid. 106, the document states that OpenAI planned to "fix" the regurgitation issue "ASAP" in order "to unlock commercialization." Microsoft defers to OpenAI.
1154. Microsoft's Director of Applied Science, Brent Hecht, expressed a concern that OpenAI would use such a filter, and called this behavior an "cover up" because it would result in "people who have a right over the content having less visibility into what was used for training."  Ex. 810 [MSFT_AI_MDL_001233834] at slide 36; Ex. 671 [Hecht Depo. Tr.] at 69:1-21.	1154. **Disputed in Part – Not Supported by Cited Evidence; Immaterial.** Microsoft responds as follows: Microsoft does not dispute that Dr. Hecht is the Director of Applied Science at Microsoft or that the quote in Paragraph 1154 reflects his testimony. But Microsoft disputes the statement in Paragraph 1154 that Dr. Hecht "expressed a concern that OpenAI would use such a filter" or that he "called this behavior an 'cover up'" because the cited evidence does not support those statements. Neither the slide nor the testimony singles out OpenAI; the slide is an industry-wide "discussion-oriented slide," and the questions concerned "companies with LLMs or products based on LLMs" and "the people engaging in those efforts" generally. News Ex. 810 (MSFT_AI_MDL_001233834) at slide 36; Ex. 212 (B. Hecht 30(b)(1) Tr.) 67:12–68:7, 69:7–21. Nor did Dr. Hecht call this behavior a "cover up." The slide refers to an "accidental coverup" and poses "Cover-up" only as a

	question, and Hecht testified that "I don't want to imply what people were intending to do" and "I don't know the intent of people working on this stuff."*Id.* Microsoft also responds that the facts set forth in Paragraph 1154 are immaterial because these facts do not affect the outcome of the suit under the governing law. Specifically, the slide poses a general discussion question about content filtering across the AI industry and does not describe what OpenAI or Microsoft actually did. News Ex. 810 (MSFT_AI_MDL_001233834) at slide 36; Ex. 212 (B. Hecht 30(b)(1) Tr.) 67:19–68:7. Additionally, Dr. Hecht's statements cannot be attributed to Microsoft. Dr. Hecht is a research academic employed at Microsoft to bring "divergent," "asymmetric, futuristic, and academic points of view." ECF 1707 (J. Usdan Decl.) ¶ 39. He does not set Microsoft's core priorities or speak for Microsoft as to "his theoretical views on AI's potential effect on content creators," and Microsoft leadership "does not share all of [his] views"; and he is a "kind of intellectual provocateur." ECF 1707 (J. Usdan Decl.) ¶ 39; Ex. 285 (J. Usdan [30(b)(6)] Tr.) 173:22-174:9. Dr. Hecht has himself disclaimed any authority to speak for Microsoft. He testified that his views on these subjects are "my view, not Microsoft's view," that "I have no idea if it's shared by the board of Microsoft," and that "I don't speak for Microsoft." Ex. 212 (B. Hecht 30(b)(1) Tr.) 319:18–20, 320:10–12, 71:5–6. He further described his views as "hypotheses," cautioned that "the jury is still out on a lot of this stuff," and acknowledged that he is "not a lawyer." *Id.* 59:11–60:8,

	165:9–15, 166:10–11. Dr. Hecht's personal views therefore are not statements of Microsoft and do not reflect Microsoft's knowledge, policies, or conduct. They have no bearing on any element of the claims or defenses asserted against Microsoft in this case. OpenAI joins Microsoft's response, except to the extent any portion of it, including any testimony quoted under Fed. R. Evid. 106, could be read to characterize OpenAI, its models, or its conduct. OpenAI disputes any such characterization. Dr. Hecht is a Microsoft researcher who testified unequivocally: "I don't speak for Microsoft. I definitely don't speak for OpenAI." Ex. 212 (B. Hecht 30(b)(1) Tr.) 100:11–12. His statements are not based on personal knowledge of OpenAI, are not attributable to OpenAI, and are inadmissible hearsay as to OpenAI. *See* Fed. R. Civ. P. 56(c)(2), (c)(4); Fed. R. Evid. 602, 801(d)(2)(D), 802. This paragraph cannot be offered to establish any fact concerning OpenAI.
1155. In June 2023, OpenAI performed an evaluation to determine the intersection between *New York Times* content and ChatGPT conversation data. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 231:15-233:6.	1155. **Undisputed.** Microsoft defers to OpenAI.
1156. The "intersection" between *New York Times* content and ChatGPT conversation data refers to instances where *New York Times* content overlaps with ChatGPT conversation data. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 231:15-233:23.	1156. **Disputed.** OpenAI responds as follows: Disputed as to News Plaintiffs' characterization of, and inference drawn therefrom, as misleading because it omits necessary context to understand the quoted language. Pursuant to Fed. R. Evid. 106, Mr. Monaco

	testified that the "intersection" measured by the eval was the presence of a matching set of n-grams, or consecutive words, between ChatGPT conversation data and a set of New York Times ngrams that could reflect, but did "[n]ot necessarily" reflect, the output of New York Times content in a ChatGPT conversation. Microsoft defers to OpenAI.
1157. OpenAI's internal documents refer to this June 2023 "intersection" analysis as "recall_pipeline/2023-06-29-nyt". Dkt. 1070 at 3, Dkt. 1070-12, Dkt. 1070-13.	1157. **Disputed in Part – Not Supported by Cited Evidence.** OpenAI responds as follows: The facts set forth in Paragraph 1157 are disputed in part because they are not supported by the cited evidence. Specifically, undisputed that there exists source code related to, but not the same as, the "intersection analysis" bearing the title "recall_pipeline.py." However, disputed that the cited evidence supports the statement. The only cited document containing the "recall_pipeline/2023-06-29-nyt" string is not an internal OpenAI document, but a brief drafted by News Plaintiffs. ECF 1070 at 3. The name "recall_pipeline/2023-06-29-nyt" does not appear in the two cited OpenAI documents. *See* ECF 1070-12; ECF 1070-13. Microsoft defers to OpenAI.
1158. In June 2023, OpenAI deployed a version of the Giraffe filter with "NYT ngrams added" in "shadow mode." Ex. 921 [OPCO_MDL_001896229] at -30; Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 377:16-378:23.	1158. **Undisputed.** OpenAI responds as follows: Undisputed that in June 2023 OpenAI had deployed an experimental version of the Giraffe bloom filter that included "NYT ngrams" to a pre-production environment

	for testing purposes and that this environment can be referred to within OpenAI as "shadow mode." Microsoft defers to OpenAI.
1159. "Shadow mode" refers to a configuration where the Giraffe filter logs instances where the output containing n-grams matching the content included in the filter but does not block ChatGPT from providing that output to the user. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 377:16-378:23.	1159. **Undisputed.** OpenAI responds as follows: Undisputed that pre-production environment used for testing, referred to internally as "shadow mode," was not a production environment and therefore versions of the Giraffe bloom filter deployed in that environment for testing would not operate block outputs from being shown to the user, though other versions of the bloom filter running in production would do so. Microsoft defers to OpenAI.
1160. When OpenAI deployed the Giraffe filter with "NYT ngrams added" in "shadow mode" in June 2023, it observed a [redacted] to the filter. Ex. 921 [OPCO_MDL_001896229] at -29.	1160. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The fact set forth in Paragraph 1160 is disputed because it is not supported by the cited evidence. Specifically, the "[redacted]" figure described in the document was the "false positive rate" or "FPR," or overlapping ngrams that do not correspond to New York Times content. News Ex. 921 at OPCO_MDL_001896229. Microsoft defers to OpenAI.
1161. OpenAI's Giraffe filter did not block outputs containing *New York Times* content until after The Times	1161. **Undisputed.** Microsoft defers to OpenAI.

sued OpenAI. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 290:6-291:7.	
1162. By December 30, 2023, OpenAI had created a Giraffe dataset comprising *New York Times* content in its training data. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 217:22-218:19.	1162. **Undisputed.** OpenAI responds as follows: Undisputed that by December 30, 2023, OpenAI had created a dataset created from records found in one or more existing training datasets that were associated with New York Times urls. Microsoft defers to OpenAI.
1163. In January 2024, OpenAI was working on constructing "a bloom filter of only nytimes data." Ex. 923 [OPCO_MDL_001805669] at -72.	1163. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The fact set forth in Paragraph 1163 is Specifically, disputed that the cited evidence supports the statement as written, as it is misleading and lacks necessary context. The cited document does not show that "OpenAI was working on constructing" such a bloom filter but instead suggests that in January 2024 an OpenAI employee considered "Figur[ing] out how to construct a bloom filter of only nytimes data" as part of an experiment, the point of which was "to verify that [a] new dataset format [was] compatible with the existing [bloom filter creation] pipeline." News Ex. 923 at OPCO_MDL_001805672. Microsoft defers to OpenAI.
1164. In April 2024, OpenAI was working on building "a smaller bloom filter w/ NYT dataset." Ex. 925 [OPCO_MDL_004775473] at -73.	1164. **Undisputed.** OpenAI responds as follows: Undisputed that in April 2024, an OpenAI employee

	observed that OpenAI was "building a smaller bloom filter w/ NYT dataset to reduce system complexity." Microsoft defers to OpenAI.
1165. The "smaller bloom filter w/ NYT dataset." reduced the size of the bloom filter from [redacted]. Ex. 925 [OPCO_MDL_004775473] at -73.	1165. **Undisputed.** OpenAI responds as follows: Undisputed that the OpenAI employee further observed that the smaller version of the bloom filter designed to "reduce system complexity" was [redacted] than the prior version. Microsoft defers to OpenAI.
1166. OpenAI referred to its process for building a smaller filter as "pruning." Ex. 927 [OPCO_MDL_004839087] at -91; Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 210:14-211:4.	1166. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that Project Giraffe involved "pruned" datasets. However, disputed as to Plaintiffs' characterizations of "pruning" as "building a smaller filter" instead of the process for removing ngrams that are highly unlikely to be regurgitated, as described in Mr. Monaco's deposition. *See* Ex. 240 (V. Monaco 30(b)(6) Tr. (Apr. 8, 2026) 215:24--217:10. The disputed proposition is not supported by the cited evidence. Microsoft defers to OpenAI.
1167. Specifically, OpenAI "pruned" the datasets it used for the Giraffe filter based on "regurgitation likelihood" of certain n-grams. Ex. 927	1167. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that Project Giraffe involved "pruned" datasets. However, disputed as to Plaintiffs' characterizations of "pruning" as

[OPCO_MDL_004839087] at -87, -91; Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 215:24-217:21].	being "based on regurgitation likelihood" instead of the process for removing ngrams that are highly unlikely to be regurgitated, as described in Mr. Monaco's deposition. *See* New Ex. 240 (V. Monaco 30(b)(6) Tr. (Apr. 8, 2026) 215:24--217:10. Microsoft defers to OpenAI.
1168. OpenAI's Giraffe pruning pipeline calculated model probabilities of n-grams as a proxy for the regurgitation likelihood of those n-grams. Ex. 927 [OPCO_MDL_004839087] at -91.	1168. **Disputed - Not Supported by Cited Evidence.** OpenAI responds as follows: The fact set forth in Paragraph 1168 is disputed because it is not supported by the cited evidence. *See* responses 1166 & 1167, *supra*. Microsoft defers to OpenAI.
1169. By May 20, 2024, OpenAI had created a Giraffe dataset comprising Daily News Plaintiffs content in its training data. Ex. 651 [Monaco Apr. 8, 2026 30(b)(6) Depo. Tr.] at 220:19-221:16.	1169. **Undisputed.** OpenAI responds as follows: Undisputed that by May 20, 2024, OpenAI had created a dataset created from records found in one or more training datasets that were associated with urls for the Daily News Plaintiffs' domains. Microsoft defers to OpenAI.
1170. OpenAI's Daily News Giraffe dataset also contains The Intercept's content. Wenger Opening Rep. at 86.	1170. **Disputed – Not Supported by Admissible Evidence.** OpenAI responds as follows: This fact is disputed because it is not supported by admissible evidence. Specifically, Dr. Wenger's *ipse dixit* conclusion is not supported by any evidence. Dr. Wenger's claim that she searched the Bloom filter for Plaintiff's domain names (e.g. "theintercept.com") is contradicted by the fact that this

	dataset does not even contain a "url" field. *See* OPCO_MDL_004658355. Microsoft defers to OpenAI.
1171. By May 24, 2024, OpenAI had built a version of the bloom filter with complete data from Daily News Plaintiffs' domains. Ex. 929 [OPCO_MDL_004731498] at -98.	1171. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Disputed to the extent that phrase "complete data" in isolation is misleading and lacks proper context. Specifically, the "complete data" as used in the document refers to records sourced from one or more of OpenAI's training datasets that can be identified (e.g., by url field) as coming from one of the Daily News Plaintiffs' domains. *See* News Ex. 929 at OPCO_MDL_004731498. Microsoft defers to OpenAI.
1172. The bloom filter with complete data from Daily News Plaintiffs' Domains was 480 GB, which OpenAI determined was too big to deploy in production. Ex. 929 [OPCO_MDL_004731498] at -98.	1172. **Disputed in Part - Not Supported by Cited Evidence.** OpenAI responds as follows: Undisputed that OpenAI created a bloom filter using data associated with the Daily News Plaintiffs' domains that was 480 GB in size and that OpenAI determined was too large to deploy in production. However, disputed to the extent that phrase "complete data" in isolation is misleading and lacks proper context. Specifically, the "complete data" as used in the document refers to records sourced from one or more of OpenAI's training datasets that can be identified (e.g., by url field) as coming from one of the Daily News Plaintiffs' domains. *See* News Ex. 929 at OPCO_MDL_004731498.

	Microsoft defers to OpenAI.
1173. To reduce the size of the bloom filter containing data from Daily News Plaintiffs Domains, OpenAI ran "model-based pruning" which determined the regurgitation likelihood of n-grams appearing in the data associated with Daily News Plaintiffs' domains. Ex. 929 [OPCO_MDL_004731498] at -98.	1173. **Disputed in Part - Not Supported by Cited Evidence**. OpenAI responds as follows: Undisputed that OpenAI ran model based pruning. However, Plaintiffs' characterization of model based pruning is not supported by the cited evidence. Specifically, disputed as to Plaintiffs' characterizations of "model-based pruning" as "determin[ing] the regurgitation likelihood of n-grams" instead of the process for removing ngrams that are highly unlikely to be regurgitated, as described in Mr. Monaco's deposition. Ex. 240 (V. Monaco 30(b)(6) Tr. (Apr. 8, 2026)) 215:24--217:10. Microsoft defers to OpenAI.
1174. OpenAI has performed at least 14 intersection evaluations using datasets containing *New York Times* and Daily News Plaintiffs' content. Ex. 931 [OPCO_MDL_004898640]; Ex. 933 [OPCO_MDL_004898659]; Ex. 935 [OPCO_MDL_004898644]; Ex. 937 [OPCO_MDL_004898645]; Ex. 939 [OPCO_MDL_004898646]; Ex. 941 [OPCO_MDL_004898647]; Ex. 943 [OPCO_MDL_004898648]; Ex. 945 [OPCO_MDL_004898649]; Ex. 947 [OPCO_MDL_004898650]; Ex. 949 [OPCO_MDL_004898651]; Ex. 951 [OPCO_MDL_004898652]; Ex. 953	1174. **Undisputed**. Microsoft defers to OpenAI.

[OPCO_MDL_004898653];	Ex.	955	
[OPCO_MDL_004898654];	Ex.	957	
[OPCO_MDL_004898655];	Ex.	959	
[OPCO_MDL_004898656];	Ex.	961	
[OPCO_MDL_004898657].			

Dated: October 23, 2026

Respectfully submitted,

*/s/ Allison S. Blanco**
LATHAM & WATKINS LLP
Andrew M. Gass (*pro hac vice*)
andrew.gass@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
sy.damle@lw.com
Luke A. Budiardjo
luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

Elana Nightingale Dawson (*pro hac vice*)
elana.nightingaledawson@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

/s/ Joseph C. Gratz
MORRISON & FOERSTER LLP
Joseph C. Gratz (*pro hac vice*)
jgratz@mofo.com
Tiffany Cheung (*pro hac vice*)
tcheung@mofo.com
Caitlin Sinclaire Blythe (*pro hac vice*)
cblythe@mofo.com
425 Market Street
San Francisco, CA 94105
Telephone: 415.268.7000

Lily Li (*pro hac vice*)
YLi@mofo.com
755 Page Mill Road
Palo Alto, CA 94304
Telephone: 650.813.5600

/s/ Robert A. Van Nest
KEKER, VAN NEST & PETERS LLP
Robert A. Van Nest (*pro hac vice*)
rvannest@keker.com
R. James Slaughter (*pro hac vice*)
rslaughter@keker.com
Paven Malhotra
pmalhotra@keker.com
Michelle S. Ybarra (*pro hac vice*)
mybarra@keker.com
Reid P. Mullen (*pro hac vice*)
rmullen@keker.com
Nicholas S. Goldberg (*pro hac vice*)
ngoldberg@keker.com
633 Battery St.
San Francisco, CA 94111
Telephone: 415.391.5400

Attorneys for Defendant OpenAI

/s/ Lisa Simpson

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)
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

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 Berkley 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