

**IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK**

ADVANCE LOCAL MEDIA LLC, ADVANCE
MAGAZINE PUBLISHERS INC. D/B/A
CONDE NAST, THE ATLANTIC MONTHLY
GROUP LLC, FORBES MEDIA LLC,
GUARDIAN NEWS & MEDIA LIMITED,
INSIDER, INC., LOS ANGELES TIMES
COMMUNICATIONS LLC, THE
MCCLATCHY COMPANY, LLC, NEWSDAY,
LLC, PLAIN DEALER PUBLISHING CO.,
POLITICO LLC, THE REPUBLICAN
COMPANY, TORONTO STAR NEWSPAPERS
LIMITED, and VOX MEDIA, LLC,

Plaintiffs,

v.

COHERE INC.,

Defendant.

Civil Action No. _____

COMPLAINT

JURY TRIAL DEMANDED

Plaintiffs Advance Local Media LLC; Advance Magazine Publishers Inc. d/b/a Condé Nast; The Atlantic Monthly Group LLC; Forbes Media LLC; Guardian News & Media Limited; Insider, Inc.; Los Angeles Times Communications LLC; The McClatchy Company, LLC; Newsday, LLC; Plain Dealer Publishing Co.; Politico LLC; The Republican Company; Toronto Star Newspapers Limited; and Vox Media, LLC (collectively “Publishers”), by their attorneys, Oppenheim + Zembrak, LLP, for their complaint against Defendant Cohere Inc. (“Cohere”), allege on personal knowledge as to matters relating to themselves and on information and belief as to all other matters, as set forth below.

NATURE OF THE CASE

1. This is a lawsuit to protect journalism from systematic copyright and trademark infringement. Rather than create its own content, Cohere takes the creative output of Publishers, some of the largest, most enduring, and most important news, magazine, and digital publishers in the United States and around the world. Without permission or compensation, Cohere uses scraped copies of our articles, through training, real-time use, and in outputs, to power its artificial intelligence (“AI”) service, which in turn competes with Publisher offerings and the emerging market for AI licensing. Not content with just stealing our works, Cohere also blatantly manufactures fake pieces and attributes them to us, misleading the public and tarnishing our brands.

2. Publishers spend enormous time and effort investigating, reporting, and ultimately publishing their expressive and groundbreaking pieces, which span the full spectrum of investigative reporting, breaking news, opinion pieces, arts and entertainment reviews, sports coverage, and political and business journalism. Collectively, Publishers reach an audience of hundreds of millions of readers in the United States alone. Their publications keep communities informed, hold those in power accountable, and support the free flow of information and ideas—functions at the core of our society and democratic system of governance. Each of the Publishers is a member of the News / Media Alliance, the leading trade association for news, magazine, and digital media organizations on matters of law and public policy.

3. Cohere is a large AI company with a valuation of over \$5 billion. Cohere freely admits that “AI is only as useful as the data it can access.”¹ Cohere also purports to distinguish

¹ Aidan Gomez (@aidangomez), X (Feb. 8, 2025, 1:48 PM), <https://x.com/aidangomez/status/1888298868171047358> (posting by Cohere founder and CEO, which was in turn reposted by Cohere (@cohere)).

itself from “Ordinary AI” that is good for “stealing intellectual property.”² Unfortunately, rather than reconcile those concepts and act lawfully, Cohere fails to license the content it uses. Cohere takes Publishers’ valuable articles, without authorization and without providing compensation. Cohere copies, uses, and disseminates Publishers’ news and magazine articles to build and deliver a commercial service that mimics, undercuts, and competes with lawful sources for their articles and that displaces existing and emerging licensing markets.

4. Cohere copies Publishers’ works to train its suite of large language model (“LLM”) AI systems, called the “Command Family” of products. Inherent to Cohere’s goal of providing trustworthy, “verifiable answers,”³ its models are designed to consult “retrievable sources” including Publishers’ websites, a feature Cohere calls “crucial for use cases like content generation” or knowledge assistance.⁴ Accessible via a Cohere interface known as “Chat,” Cohere’s AI models deliver outputs that include full verbatim copies, substantial excerpts, and substitutive summaries of Publishers’ works—even current, breaking news pieces and articles protected by paywalls. Cohere also delivers material that it represents came from Publishers but in fact is material that Cohere produced. This retrieval-augmented generation feature, where an LLM’s output draws on not just what the LLM trained on, but also other sources such as websites, is referred to as “RAG,” a term coined by a Cohere researcher.

5. Publishers value innovation and the promise that artificial intelligence holds if ethically deployed. In fact, many Publishers already license their articles to AI companies. But Cohere improperly usurps Publishers’ creative labor and investments for the sake of its own profits.

² *Id.*

³ Aidan Gomez, *Where Enterprise AI is Headed*, LINKEDIN (Dec. 5, 2024), <https://www.linkedin.com/pulse/where-enterprise-ai-headed-aidan-gomez-dznjf>.

⁴ *Introducing Chat with Retrieval-Augmented Generation (RAG)*, COHERE (Sept. 28, 2023), <https://cohere.com/blog/chat-with-rag>.

6. Publicly, Cohere claims to “care a lot about the provenance of our data,” including “whether we have permission to train on that data,” and has indicated that “we don’t scrape stuff that we shouldn’t scrape.”⁵ Cohere also claims that it does not want “to be training on stuff that people don’t want us training on, full stop.”⁶ That is false.

7. Cohere helps itself to unlicensed copies of Publishers’ news and magazine articles to build a training dataset that it supplements with further real-time copying of articles, infringing Publishers’ copyrights and seizing an unfair competitive advantage against AI companies that abide by the law. Cohere further infringes Publishers’ copyrights by providing copies of Publishers’ articles, undermining lawful sources for Publishers’ content. Cohere also delivers fake pieces under Publishers’ names, confusing the public and damaging Publishers’ brands by falsely associating Publishers with content that lacks the quality that readers expect from Publishers.

8. Cohere’s actions amount to massive, systematic copyright infringement and trademark infringement, and have caused significant injury to Publishers. Left unfettered, such misconduct threatens the continued availability of the valuable news, magazine, and media content that Publishers produce. Indeed, Cohere admits that it seeks to supplant publishers, luring readers with a promise that its models can “keep you up to date with the latest news.”⁷

9. Cohere must be enjoined and held to account for the harm it causes. Publishers’ Complaint includes the allegations set forth in these numbered paragraphs, as well as a series of exhibits. Exhibit A, which identifies over four thousand articles, is a non-exhaustive, illustrative

⁵ Sharon Goldman, *Cohere CEO and president discuss Nvidia and Oracle funding, Hinton comments and the future of LLMs*, VENTUREBEAT (June 27, 2023), <https://venturebeat.com/ai/cohere-ceo-and-president-discuss-nvidia-and-oracle-funding-hinton-comments-and-the-future-of-llms>.

⁶ Nilay Patel, *AI will make money sooner than you’d think says Cohere CEO Aidan Gomez*, THE VERGE (June 10, 2024), <https://www.theverge.com/24173858/ai-cohere-aidan-gomez-money-revenue-llm-transformers-enterprise-stochastic-parrot>.

⁷ *The Cohere AI App is available in Slack*, COHERE (Sept. 16, 2024), <https://cohere.com/blog/slack-cohere-ai-app>; *accord Introducing Chat with Retrieval-Augmented Generation (RAG)*, COHERE (Sept. 28, 2023), <https://cohere.com/blog/chat-with-rag>.

list of Publishers’ copyrighted works that Cohere has infringed. Exhibit B contains some specific examples of copyright-infringing outputs that Cohere provides. Exhibit C is a non-exhaustive, illustrative list of Publishers’ federally registered trademarks that Cohere has infringed. Exhibit D contains examples of misleading outputs where Cohere passes off its own hallucinated articles as articles from Publishers.

10. The breadth of Cohere’s misconduct is extensive. While some of it is visible, much of it occurs behind closed doors. Publishers reserve the right to amend this Complaint as the case proceeds, including based on information obtained in discovery.

THE PARTIES

Advance Companies

11. Plaintiff Advance Magazine Publishers Inc. (“Condé Nast”) is a New York corporation with its principal place of business in New York, New York. Founded in 1909 by publisher Condé Montrose Nast, Condé Nast is home to some of the most iconic magazine brands in media, including *Vogue*, *The New Yorker*, *GQ*, *Vanity Fair*, *Wired*, *Bon Appétit*, and *Architectural Digest*. Condé Nast creates and distributes media across all formats, including print, video, film, digital, audio, and social media channels. In the last five years, Condé Nast publications have won four Pulitzer Prizes, including in 2024 for *The New Yorker*’s explanatory reporting on the social consequences of the felony murder doctrine. Exhibit A includes a representative sample of Condé Nast’s works—as published in *The New Yorker*, *Vogue*, *Architectural Digest*, and *Wired*—that Cohere has infringed.

12. Plaintiff Advance Local Media LLC (“Advance Local Media”) is a New York corporation with its principal place of business in New York, New York. Advance, named after the *Staten Island Advance*, dates back to 1922 and is known for high-quality local journalism.

Fulfilling its mission, “To strengthen and empower the communities we serve,” Advance’s publications and media groups reach 52 million people a month. It has won four Pulitzer Prizes in the past five years, including two in 2023 for its Alabama newspaper’s investigation of predatory policing in Brookside, Alabama and AL.com columnist Kyle Whitmire’s “State of Denial” series, exploring how Confederate heritage continues to shape the state. Advance publications are available online, and many are also in print. Exhibit A includes a representative sample of Advance’s works, as published in *The Oregonian*, that Cohere has infringed.

13. Plaintiff Plain Dealer Publishing Co. (“The Plain Dealer”) is an Ohio corporation with a principal place of business in Brooklyn, Ohio, and publisher of *The Plain Dealer*. Continuously published since 1842, *The Plain Dealer* is Ohio’s largest newspaper by print circulation and has established a reputation for excellence in local journalism. *The Plain Dealer* has won numerous awards, including the 2005 Pulitzer Prize in Commentary. Exhibit A includes a representative sample of The Plain Dealer’s works that Cohere has infringed.

14. Plaintiff The Republican Company (“The Republican”), publisher of *The Republican*, is a Massachusetts corporation with its principal place of business in Springfield, Massachusetts. Founded as a weekly newspaper in 1824, *The Republican* has long been known for the quality of its reporting. In 2021, *The Republican* was recognized for the fourth consecutive year as “Newspaper of the Year” by the New England Newspaper Association. Exhibit A includes a representative sample of The Republican’s works that Cohere has infringed.

15. Advance Publications, Inc. is a New York-based privately held company that operates and owns a broad range of media and communications businesses, including local news media companies that produce newspapers across ten states. Condé Nast, Advance Local Media,

The Plain Dealer, and The Republican are subsidiaries, directly or indirectly, of Advance Publications, Inc.

16. Advance Local Media, The Plain Dealer, and The Republican and are collectively referred to herein as “Advance.”

The Atlantic

17. Plaintiff The Atlantic Monthly Group LLC d/b/a The Atlantic (“The Atlantic”) is a Delaware corporation with its principal place of business in Washington, D.C. The Atlantic also maintains a significant corporate office in New York City. Founded in 1857 in Boston as *The Atlantic Monthly*, *The Atlantic* magazine began by publishing leading commentary on opposition to slavery, alongside coverage of literature, politics, and arts, from great American writers, including Ralph Waldo Emerson, Oliver Wendell Holmes, Harriet Beecher Stowe, and Henry Wadsworth Longfellow. It continues that tradition today with important coverage of consequential political and social issues. In 2024, for the third consecutive year, *The Atlantic* won the top honor of General Excellence for a News, Sports, and Entertainment publication at the 2024 National Magazine Awards, among many accolades. *The Atlantic* is offered in print and online. Exhibit A includes a representative sample of The Atlantic’s works that Cohere has infringed.

Axel Springer Companies

18. Plaintiff Insider, Inc. (“Business Insider”), formerly known as Business Insider Inc., is a Delaware corporation with its principal place of business in New York, New York. Business Insider publishes *Business Insider*, a popular news website that has offered the latest updates on business, financial, political, sports, lifestyle, and technology since 2007. *Business Insider* has received numerous accolades for its dynamic original reporting, including a 2024 Excellence in Financial Journalism Award for a piece on artificial intelligence and the workforce, a 2024 Hillman

Award in Newspaper Journalism for “harrowing, eye-opening reporting” on the use of trained attack dogs in Virginia prisons, and a 2022 Pulitzer Prize in Illustrated Reporting and Commentary for the piece *How I escaped a Chinese internment camp*. Exhibit A includes a representative sample of Business Insider’s works that Cohere has infringed.

19. Plaintiff Politico LLC (“Politico”) is a Delaware corporation with a principal place of business in Arlington, Virginia and the publisher of the online news outlet *Politico*. Founded in 2007, *Politico* covers fast-breaking politics and policy in the United States and abroad. *Politico*’s journalism has won acclaim across a range of categories, including a 2023 George Polk Award for National Reporting, a 2022 James Beard Award nomination for Health and Wellness Writing, and a 2014 Gerald R. Ford Journalism Prize for Distinguished Reporting on the Presidency. Exhibit A includes a representative sample of Politico’s works that Cohere has infringed.

20. Axel Springer SE (“Axel Springer”) is a German company that publishes dozens of newspapers and magazines, both online and in print, and owns the English-language publications *Business Insider* and *Politico*. Business Insider and Politico are collectively referred to herein as “Axel Springer Companies.”

Forbes

21. Plaintiff Forbes Media LLC (“Forbes”) is a Delaware corporation with its principal place of business in Jersey City, New Jersey. Founded in 1917, it publishes the magazine and website *Forbes*, featuring articles on business, investing, technology, entrepreneurship, science, communications, and lifestyle topics. In addition to its reporting on finance and business, *Forbes* is known for its lists and rankings, such as “The World’s Most Powerful People” and “30 Under 30.” *Forbes* is published in print and online. Exhibit A includes a representative sample of Forbes’ works that Cohere has infringed.

The Guardian

22. Plaintiff Guardian News & Media Limited (“The Guardian”), publisher of *The Guardian*, is a United Kingdom corporation with its principal place of business in London, England. Founded in 1821, *The Guardian* is a United Kingdom newspaper of record and launched its US-based online edition in 2011. The headquarters for its U.S. edition is in New York City. Guardian US is known for investigative journalism and balanced reporting including pieces on National Security Agency revelations, the Panama Papers document leak, and a 2015-2016 investigation into the number of people killed by law enforcement titled “The Counted.” *The Guardian* has received numerous awards, including a Pulitzer Prize in 2014 for reporting on widespread secret surveillance by the National Security Agency. Exhibit A includes a representative sample of The Guardian’s works that Cohere has infringed.

LA Times

23. Plaintiff Los Angeles Times Communications LLC (“LA Times”), publisher of the *Los Angeles Times*, is a California corporation with a principal place of business in Los Angeles, California. Founded as the *Los Angeles Daily Times* in 1881, the *Los Angeles Times* is one of the largest daily newspapers in the United States. It is published in print and online and its website, latimes.com, receives over 40 million unique visitors each month. From 1942 to today, it has won 52 Pulitzer Prizes, including six gold medals for public service. These Pulitzer Prizes include awards for spot news reporting for the 1965 Watts Riots and 1992 Los Angeles Riots, 2021 Editorial Writing for Robert Greene pieces on policing and bail reform, and in 2024, an award for Justin Chang for film criticism. Exhibit A includes a representative sample of LA Times’s works that Cohere has infringed.

McClatchy

24. Plaintiff The McClatchy Company, LLC (“McClatchy”) is a Delaware corporation with its principal place of business in Chatham, New Jersey. Founded in 1857, McClatchy is one of largest newspaper publishers in the United States, publishing 30 daily local newspapers nationwide. These papers cover local, national, and global news and have over 65 million monthly readers. McClatchy’s dedication to quality local journalism is reflected by 57 Pulitzer Prizes across McClatchy’s publications, including a 2022 award to the *Miami Herald* for its urgent, thorough, and compassionate coverage of the deadly collapse of a beachfront condominium in Surfside, Florida. Exhibit A includes a representative sample of McClatchy’s works, as published in the *Miami Herald*, the *Idaho Statesman*, and the *Fort Worth Star-Telegram*, that Cohere has infringed.

Newsday

25. Plaintiff Newsday, LLC (“Newsday”) is a Delaware limited liability company with its principal place of business in Melville, New York and the publisher of *Newsday*. Founded in 1940, *Newsday* is a daily newspaper sold throughout the New York City metropolitan area, with a focus on Nassau and Suffolk counties. It is published in print and online (including an e-edition), and serves its audience through its portfolio of mobile, digital, video, audio, print products, and live events. As of 2023, *Newsday* was the eighth-largest newspaper in the United States by circulation. *Newsday* has garnered numerous awards for its local news, spot reporting, photography, sports, and commentary, including 19 Pulitzer Prizes and 39 Emmy Awards. Exhibit A includes a representative sample of Newsday’s works that Cohere has infringed.

Torstar

26. Plaintiff Toronto Star Newspapers Limited (“Torstar”), publisher of the *Toronto Star*, is a company incorporated under the laws of Ontario with its principal place of business in

Toronto, Canada. Established in 1892, the *Toronto Star* is Canada's largest daily newspaper. It is known for the depth of its reporting on local, national, and international issues, its award-winning investigations team and being a champion of social and economic reform. Torstar and its journalists have received numerous awards including 158 National Newspaper Awards, 14 Canadian Association of Journalists investigative journalism awards, two Canadian Hillman Prizes, a Pulitzer Prize, and six Michener Awards. The *Toronto Star* is available in print and online, including digitally at thestar.com, on Torstar's mobile application, and in e-paper format. Exhibit A includes a representative sample of Torstar's works that Cohere has infringed.

Vox Media

73. Plaintiff Vox Media, LLC ("Vox Media") is a Delaware limited liability company with its principal place of business in Washington, D.C and largest office in New York, New York. Founded in 2011, Vox Media is a digital and print powerhouse, home to influential and cutting-edge media brands *The Verge*, *SB Nation*, *Vox*, and *New York Magazine*, including the sites *Intelligencer* (politics, business, sports, and technology news), *The Cut* (fashion), *Vulture* (arts and entertainment), *The Strategist* (product reviews and recommendations), *Curbed* (real estate, architecture, and design), and *Grub Street* (food). Vox Media recently won its first Pulitzer Prize when *New York Magazine*'s book critic, Andrea Long Chu, was awarded the 2023 Pulitzer Prize in Criticism. Exhibit A includes a representative sample of Vox Media's works, as published in *Vox*, *Vulture*, *Grub Street*, and *New York Magazine*, that Cohere has infringed.

Cohere

27. Defendant Cohere is a Canadian company with its principal places of business at 171 John St 2nd Floor, Toronto, Ontario, Canada and 394 Pacific Avenue, San Francisco,

California. Cohere maintains an office at 1 Little West 12th Street, New York, New York with approximately 30 employees.

JURISDICTION AND VENUE

28. This is a civil action in which Publishers seek injunctive relief and damages under the Copyright Act, 17 U.S.C. § 101, *et seq.*, and under the Lanham Act, 15 U.S.C. § 1051, *et seq.* The Court has original subject matter jurisdiction over Publishers’ copyright and trademark claims pursuant to 28 U.S.C. §§ 1331 and 1338(a).

29. This Court has specific personal jurisdiction over Cohere pursuant to N.Y. C.P.L.R. §§ 302(a)(1), (3)–(4).

30. Cohere knowingly and intentionally markets and supplies its services to persons in New York and this District. Cohere maintains a New York City office, which, as Cohere explains on its website, “serve[s] as an important hub for its senior leadership team to connect with its many existing partners in the city.” As Cohere CEO Aidan Gomez has added, “New York is a critical talent market for [Cohere],” and accordingly Cohere employs personnel in this District. Cohere’s New York City–based personnel are involved in creating, maintaining, and monetizing Cohere’s products, including the Command Family of products. In addition, Cohere has solicited customers and developers at a variety of in-person events in New York City. Cohere has also entered partnerships with consultancies and technology providers, including NYC-based businesses like Accenture and LivePerson, to reach a wider customer base and assist them in operating and accessing Cohere models. Further, Cohere directs its infringing activities at New York. Cohere unlawfully distributes and publicly displays the violative outputs via its Chat interface to the Command Family of LLMs to users in New York. The activity for which Cohere is liable requires a high degree of interaction between Cohere’s servers and users’ computers or mobile devices.

Cohere also enters into agreements with licensees of the Command Family in New York and derives substantial revenue from interstate and international commerce.

31. Publishers' claims arise from and relate to Cohere's activities in and contacts with this District, including Cohere providing the violative output complained of herein to requests from users and computer servers in New York. Cohere's conduct causes Publishers to suffer harm in New York. As described below, Publishers have their principal place of business, business operations, and/or customers in New York.

32. Venue is proper in this District under 28 U.S.C. §§ 1391(b) and 1400(a) because Cohere maintains an office in New York; conducts, transacts, and/or solicits business in this District; a substantial portion of the events or omissions giving rise to the asserted claims, including those described above, occurred in this District; and much of the harm caused by Cohere is felt in this District.

FACTUAL ALLEGATIONS

I. Publishers and Their Businesses

33. Publishers are globally recognized news, magazine, and digital publishers. Their publications collectively report local, national, and global news, covering science, health, business, economics, sports, technology, and culture, and providing information, analysis, and commentary to millions of people. Combined, Publishers employ tens of thousands of employees for their media outlets. Their pieces are the work of not just the journalists who research and write them, but editors and fact checkers to ensure accuracy and fairness, along with legal, operational, security, design, and other support. The distribution of news also requires marketing, advertising, subscription, and other staff to get the pieces in the hands of the public. Their copyrighted works are thus the product of enormous expertise, talent, research, investment, courage, and grit.

A. The Markets for Publishers' Works

34. While the strong public interest inherent in news publishing drives many Publishers to ensure their content reaches human readers of various economic means, Publishers are companies that need income to fund the people and resources that enable them to fulfill that public duty. Publishers' works are disseminated to the public through a variety of income-generating distribution methods, typically through a mix of subscription or ad-supported models via print or online versions, and by selling or licensing individual pieces à la carte.

35. Publishers also license their content to other businesses for many different uses. For example, many Publishers license their digital editions to aggregation platforms such as Amazon, Apple News, Microsoft News, Yahoo!, Smart News, or NewsBreak, to clippings or media monitoring services, to various social media platforms, to research outlets or intelligence platforms like Ancestry.com, Factiva, ProQuest, or Lexis Nexis, or to other domain-specific partners, including educational, library, marketing, travel, or corporate-oriented uses, such as through the Copyright Clearance Center. To facilitate the delivery of news content to readers across communities, Publishers participate in cooperative licensing agreements with organizations such as the Associated Press. Publishers also negotiate reprint or derivative licenses on an aggregate or one-off basis for different uses, either directly or through licensing intermediaries. Publishers expend significant resources in developing and pursuing new licensing opportunities and business models for their works. These licensing initiatives are an important and necessary source of income for many publishers.

36. Publishers have embraced technological innovation and provide their works to an array of online businesses. In recent years as AI models have become more prevalent, a new market has emerged for news, magazine, and digital publishers to license their content to AI developers

to use in various ways. Many Publishers, including Vox Media, The Atlantic, Condé Nast, Guardian, Los Angeles Times, Politico, and Business Insider have executed content licenses with various AI companies.

37. In December 2023, Axel Springer announced “a global partnership” with OpenAI, allowing OpenAI to “use quality content from Axel Springer media brands for advancing the training of OpenAI’s sophisticated large language models.”⁸ As part of the OpenAI–Axel Springer agreement, “ChatGPT users around the world will receive summaries of selected global news content from Axel Springer’s media brands including POLITICO [and] BUSINESS INSIDER. . . including otherwise paid content,” on the condition that “user queries will include attribution and links to the full articles for transparency and further information.”⁹ In May 2024, *The Atlantic* announced that it had licensed its articles to be “discoverable within OpenAI’s products, including ChatGPT” and that under this agreement, “[q]ueries that surface *The Atlantic* will include attribution and a link to read the full article on theatlantic.com.”¹⁰ Vox Media issued a similar announcement that same day.¹¹ Likewise, in August 2024, Condé Nast and OpenAI announced that Condé Nast had concluded an agreement with the artificial intelligence company “to display content from top brands like Vogue, The New Yorker, Condé Nast Traveler, GQ, Architectural

⁸ *Axel Springer and OpenAI partner to deepen beneficial use of AI in journalism*, AXEL SPRINGER (Dec. 13, 2023), <https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism>.

⁹ *Id.*

¹⁰ *The Atlantic announces product and content partnership with OpenAI*, THE ATLANTIC (May 29, 2024), <https://www.theatlantic.com/press-releases/archive/2024/05/atlantic-product-content-partnership-openai/678529>; see also Nilay Patel, *Why The Atlantic signed a deal with OpenAI*, THE VERGE (July 11, 2024, 1:05 pm EDT), <https://www.theverge.com/2024/7/11/24196396/the-atlantic-openai-licensing-deal-ai-news-journalism-web-future-decoder-podcasts> (explaining that the CEO of *The Atlantic* “just signed a deal allowing OpenAI to use *The Atlantic*’s vast archives as training data”).

¹¹ Sara Fischer, *Exclusive: The Atlantic, Vox Media ink licensing, product deals with OpenAI*, AXIOS (May 29, 2024), <https://www.axios.com/2024/05/29/atlantic-vox-media-openai-licensing-deal>.

Digest, Vanity Fair, Wired, Bon Appetit, and more, within [OpenAI’s] products, including ChatGPT and our SearchGPT prototype,” with “direct links to news stories.”¹²

38. These deals are not limited to a single AI company. For instance, several Publishers have entered into arrangements with ProRata.ai.¹³ Recognizing that LLMs rely on news content to “ensure[] that [thei]r responses to user queries remain comprehensive (and) nuanced,” the AI company Perplexity reached an agreement with dozens of news organizations, including LA Times, to “shar[e] ad revenue from interactions” with Perplexity’s LLM “when a publisher’s content is referenced.”¹⁴ In addition, the Copyright Clearance Center’s Annual Copyright License authorizes participating businesses to use news publisher content in internal AI systems.

39. Non-plaintiff news, magazine, and digital publishers have inked similar deals. Dow Jones alone has entered into AI licensing agreements with more than 4,000 news outlets to use their articles to generate AI news summaries in its Factiva platform.¹⁵ In July 2023, the Associated Press reported that it had licensed its archive of articles to OpenAI for use as training data;¹⁶ it announced a similar deal with Google in January 2025.¹⁷ Others with licensing deals with OpenAI include DotDash Meredith, publisher of dozens of titles ranging from *People Magazine* to

¹² *OpenAI partners with Condé Nast*, OPENAI (Aug. 20, 2024), <https://openai.com/index/conde-nast>; see also Katie Knibbs, *Condé Nast Signs Deal with OpenAI*, WIRED (Aug. 20, 2024, 2:00 pm), <https://www.wired.com/story/conde-nast-openai-deal>.

¹³ Trishla Ostwal, *Exclusive: Startup ProRata.ai Launches AI Search Engine to Pay Content Creators*, ADWEEK (Dec. 9, 2024), <https://www.adweek.com/media/prorata-search-engine>; accord Bron Maher, *DMG Media invests in publisher-friendly generative AI start-up Prorata*, PRESS GAZETTE (Nov. 21, 2024), <https://pressgazette.co.uk/platforms/prorata-ai-dmg-media-guardian-sky-news>.

¹⁴ *AI startup Perplexity adds The Independent, LA Times to its publishers’ program*, REUTERS (Dec. 5, 2024, 12:10 pm EST), <https://www.reuters.com/business/media-telecom/ai-startup-perplexity-adds-independent-la-times-its-publishers-program-2024-12-05>.

¹⁵ Trishla Ostwal, *Dow Jones Wins AI Licensing Deals With More Than 4,000 News Outlets Like the AP*, ADWEEK (Nov. 12, 2024), <https://www.adweek.com/media/dow-jones-lands-more-than-4000-ai-licensing-deals-with-news-outlets-like-the-ap>.

¹⁶ Matt O’Brien, *ChatGPT-maker OpenAI signs deal with AP to license news stories*, AP NEWS (July 13, 2023), <https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a>.

¹⁷ Kyle Wiggers, *Google inks deal with The Associated Press to bring more real-time info to Gemini*, TECHCRUNCH (Jan. 15, 2025, 8:57 AM PST), <https://techcrunch.com/2025/01/15/google-inks-deal-with-the-associated-press-to-bring-more-real-time-info-to-gemini>.

Investopedia;¹⁸ leading magazine and newspaper publisher Hearst;¹⁹ News Corp, publisher of the *Wall Street Journal*, *Barron's*, and *New York Post*, among other titles;²⁰ and Future, publisher of *TechRadar*, *Tom's Guide*, and *Marie Claire*.²¹ In October 2024, Reuters entered a multiyear licensing agreement allowing Meta to “use Reuters content to answer user questions in real time about news and current events” via its artificial intelligence chatbot Meta AI, including by providing “summaries and links to Reuters content.”²² In January 2025, Agence France-Presse entered into a deal with AI company Mistral to “improve the accuracy of answers in Le Chat, Mistral’s chatbot,” which Mistral’s CEO called “a key step in the deployment of our technology, particularly for businesses.”²³ The prevalence of these agreements show the strength of the market for news content to train LLMs, for copying to provide real-time information, and to display summaries in the output of LLMs.

B. Publishers’ Copyrights

40. Publishers expend resources and employ legal and technical means to protect their valuable content, ranging from registration, copyright notices, and terms of service of their sites restricting use to access and anti-copying controls, such as paywalls or metering systems, or machine-readable protocols like robots.txt. Publishers require third party businesses to agree to

¹⁸ *Investopedia-owner Dotdash Meredith signs content license deal with OpenAI*, REUTERS (May 7, 2024, 9:23 am EDT), <https://www.reuters.com/markets/deals/investopedia-owner-dotdash-meredith-signs-content-license-deal-with-openai-2024-05-07>.

¹⁹ Winston Cho, *OpenAI Inks Deal with Hearst, Marking Another Major Media Partnership*, THE HOLLYWOOD REPORTER (Oct. 8, 2024), <https://www.hollywoodreporter.com/business/business-news/openai-hearst-1236026980>.

²⁰ Alexandra Bruell, *OpenAI, WSJ Owner News Corp Strike Content Deal Valued at Over \$250 Million*, WALL STREET JOURNAL (May 22, 2024, 9:45 pm EDT), <https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba>.

²¹ Emma Roth, *OpenAI strikes content deal with Tom's Guide owner Future*, THE VERGE (Dec. 5, 2024, 8:41 am EST), <https://www.theverge.com/2024/12/5/24313909/openai-content-deal-toms-guide-future>.

²² *Meta Platforms to use Reuters news content in AI chatbot*, REUTERS (Oct. 25, 2024, 4:54 pm EDT), <https://www.reuters.com/technology/artificial-intelligence/meta-platforms-use-reuters-news-content-ai-chatbot-2024-10-25>.

²³ Romain Dillet, *Mistral signs deal with AFP to offer up-to-date answers in Le Chat*, TECHCRUNCH (Jan. 16, 2025), <https://techcrunch.com/2025/01/16/mistral-signs-deal-with-afp-to-offer-up-to-date-answers-in-le-chat>.

licenses that specify the terms on which business may use, reproduce, and/or display Publisher content and the required payment.

41. Publishers own and/or control the exclusive rights set forth in 17 U.S.C. § 106 of the Copyright Act, in whole or in part, with respect to millions of valuable works, including the works listed in Exhibit A (the “Works in Suit”), which is an illustrative and non-exhaustive list of works that Cohere has infringed. All the works listed on Exhibit A constitute original works and copyrightable subject matter pursuant to the Copyright Act, 17 U.S.C. § 101, *et seq.*, and they have been duly registered with the U.S. Copyright Office. The copyrights in the works set forth in Exhibit A remain valid and subsisting and at all times relevant to the allegations in this Complaint have been owned and/or controlled by Publishers.

C. Publishers’ Trademarks

42. Publishers distribute their works to the public using the names of their publications, valuable trademarks denoting the source of the pieces.

43. Advance Local Media is the owner of the valid, protectable trademark THE OREGONIAN. Advance Local Media and its predecessors-in-interest have used the mark THE OREGONIAN to sell and promote newspapers since at least December 6, 1931. *The Oregonian* has earned some of journalism’s top honors, including a 2001 Pulitzer Prize in Public Service “[f]or its detailed and unflinching examination of systematic problems within the U.S. Immigration and Naturalization Service.”²⁴ Because of *The Oregonian*’s rigorous reporting, recognized contributions to journalism, and reputation for accuracy, the mark THE OREGONIAN used in connection with news represents significant consumer goodwill, developed over nearly a century.

²⁴ See 2001 Pulitzer Prizes: Journalism, THE PULITZER PRIZES, <https://www.pulitzer.org/prize-winners-by-year/2001>.

44. The Plain Dealer is the owner of the valid, protectable trademark THE PLAIN DEALER, which it has continuously and exclusively used in connection with newspapers and magazines since at least September 16, 1961. Founded as *The Cleveland Plain Dealer* in 1842, the newspaper has long been known as simply “The Plain Dealer,” prompting Winston Churchill to remark during a lecture that “by all odds the Plain Dealer has the best newspaper name of any in the world.”²⁵ Known for its timely, accurate reporting on both local and national news, *The Plain Dealer* and its staff have received some of the highest plaudits in its field, including the 2005 Pulitzer Prize in Commentary and *Editor & Publisher’s* 2003 Editor of the Year Award. Accordingly, the trademark THE PLAIN DEALER enjoys a reputation for astute reporting in the eyes of its subscribers and the public.

45. The Republican and its predecessors-in-interest have exclusively used the mark THE REPUBLICAN in connection with the publication of news articles since at least as early as 2003. As a result of this exclusive and uninterrupted use, consumers recognize that the mark THE REPUBLICAN when used in connection with newspapers identifies The Republican’s goods and services. *The Republican* has been lauded for the quality of its reporting, winning the New England Newspaper & Press Association’s top honors for “General Excellence” among daily newspapers in 2021 and the “Newspaper of the Year” award for large-circulation daily newspapers in 2019. This reputation for excellence has created significant goodwill in the mark THE REPUBLICAN. The Republican has used and promoted the mark THE REPUBLICAN over decades, thereby developing developed strong common-law trademark rights in THE REPUBLICAN.

46. Condé Nast is the owner of several valid, protectable marks, including the trademarks VOGUE, THE NEW YORKER, ARCHITECTURAL DIGEST, and WIRED

²⁵ *The Press: Cleveland Centenarian*, TIME MAGAZINE (Jan. 12, 1942), <https://content.time.com/time/subscriber/article/0,33009,773027,00.html>.

(collectively, the “Condé Nast Marks”). Condé Nast and its predecessor-in-interest, Condé Nast Publications Inc., have used the trademark VOGUE in connection with magazine articles since the first issue of *Vogue* was published on December 17, 1892. After 133 years of continuous and exclusive use, the VOGUE mark has amassed unparalleled consumer recognition and goodwill, leading commentators to call *Vogue* “the world’s most famous fashion magazine.”²⁶ Condé Nast’s rights in the trademark THE NEW YORKER are similarly longstanding. Condé Nast and its predecessor-interest, The New Yorker Magazine, Inc., have used the trademark THE NEW YORKER in connection with magazine articles since the first issue of *The New Yorker* was published on February 21, 1925. *The New Yorker* has demonstrated a commitment to excellence and earned a reputation for accuracy—so much so that “no publication has been more consistently identified with its rigorous fact-checking than *The New Yorker*.”²⁷ In the ten years since online and print magazines became eligible, *The New Yorker* has earned eight Pulitzer Prizes, including the gold medal for Public Service in 2018. Condé Nast and its predecessor, Architectural Digest Publishing Corporation, have used the ARCHITECTURAL DIGEST trademark since the first issue of *Architectural Digest* was published in 1920. Through a century of continuous promotion, Condé Nast has established *Architectural Digest* as a leading authority on architecture, design, home décor, lifestyle, and culture, leading consumers to associate the ARCHITECTURAL DIGEST mark with quality and expertise. Condé Nast and its predecessor-in-interest, Wired Ventures, Inc., have used the trademark WIRED in connection with technology-focused magazine articles since at least as early as April 1991. The recipient of five National Magazine Awards for General Excellence, *Wired* has quickly become recognized among the consuming public as an

²⁶ NORBERTO ANGELETTI, IN *VOGUE: AN ILLUSTRATED HISTORY OF THE WORLD’S MOST FAMOUS FASHION MAGAZINE* (2012).

²⁷ Colin Dickey, *The Rise and Fall of Facts*, COLUM. J. REV. (Dec. 6, 2019), https://www.cjr.org/special_report/rise-and-fall-of-fact-checking.php.

authority on technological matters, and the WIRED brand is therefore associated with trustworthy journalism.

47. The Atlantic is the owner of the valid, protectable trademark THE ATLANTIC, which it has used in connection with the publication of general interest news stories, features, and political commentary since 1857. After over 150 years of extensive use and promotion, *The Atlantic* has become a household name, associated with award-winning feature writing and in-depth reporting. In 2024 alone, *The Atlantic* won National Magazine Awards for Profile Writing, Reviews and Criticism, Columns & Essays, and the organization's highest honor, the National Magazine Award for General Excellence. In addition to its critical acclaim, *The Atlantic* enjoys considerable commercial success, with over one million paid subscriptions.

48. Insider, Inc., formerly known as Business Insider, Inc., is the owner of valid, protectable rights in the BUSINESS INSIDER mark. Business Insider has used the BUSINESS INSIDER mark in connection with providing online news about business, technology, innovation, and beyond since at least as early as 2009. *Business Insider* has developed a brand as a trusted source of original reporting, including profiles of the visionaries who built some of today's largest technology companies. In recent years, *Business Insider* has received journalism's highest honors, including a Pulitzer Prize for Illustrated Reporting and Commentary in 2022.²⁸ Through Business Insider's extensive use and promotion of the BUSINESS INSIDER mark, the trademark has become a symbol of timely, credible news and therefore represents significant goodwill.

49. Politico is the publisher of the online news publication *Politico* and the owner of the valid, protectable POLITICO trademark. Politico and its predecessors-in-interest have used the POLITICO trademark in connection with political news reporting since at least as early as 2007.

²⁸ See *Business Insider's Awards and Accolades*, BUSINESS INSIDER (Aug. 14, 2024, 12:35 pm ET), <https://www.businessinsider.com/awards>.

In addition to its considerable acclaim for its journalism, *Politico* has gained recognition within the industry and among the public as a global nonpartisan politics and policy news organization. *Politico*'s reporting and professional subscription services (such as POLITICO Pro) connect the dots between global centers of political influence. Its journalism lives online at POLITICO.com and politico.eu; with its POLITICO Pro Platforms in the U.S. and Europe; in daily and weekly newsletters, such as its POLITICO Playbook franchise, which has editions in 13 cities; and through live events.

50. Forbes is the owner of several valid, protectable trademarks. Forbes has used the FORBES mark in connection with news and magazine articles since the first issue of *Forbes* magazine was published in September 1917. *Forbes* quickly distinguished itself as a leader in business reporting and analysis by publishing influential annual report cards assessing the strengths of investments, companies, and industries. *Forbes* innovated how the media reports on wealth and success, developing and producing widely circulated and highly regarded lists and rankings, including *Forbes*' 30 Under 30 Lists and The *Forbes* 400. Today, *Forbes* is available in print and online through forbes.com. Its digital footprint is sizeable, with over 150 million monthly visitors to *Forbes*-branded websites and social media and over 1 million subscribers to its daily email newsletters. Through Forbes' consistent use and promotion over decades, the FORBES mark has become synonymous with quality reporting and analysis on business and finance.

51. The Guardian is the owner of the valid, protectable trademark THE GUARDIAN. The Guardian and its predecessors-in-interest have used the trademark THE GUARDIAN in connection with the sale and promotion of newspaper articles in the United States since at least 1959 and in connection with the online delivery of newspaper articles since at least 2011. Founded as *The Manchester Guardian* in 1821 and branded as simply *The Guardian* since 1959, *The*

Guardian has long been regarded as a newspaper of record in the United Kingdom and a respected news source in the United States. *The Guardian US* has earned accolades in the United States for its journalism, including a 2022 Online Journalism Award for its investigative reporting and in-depth analysis of an Israeli surveillance company and, that same year, the National Press Club Sandy Hume Memorial Award for Excellence in Political Journalism for *The Guardian's* coverage of U.S. politics. *The Guardian* is a global brand known for detailed reporting on a broad range of topics, and U.S. consumers have come to recognize the trademark THE GUARDIAN as exclusively identifying *The Guardian's* news reporting and The Guardian's goods and services.

52. LA Times is the owner of several valid, protectable marks including the trademarks LOS ANGELES TIMES and LATIMES.COM (collectively, the "LA Times Marks"). LA Times and its predecessors-in-interest have used the LOS ANGELES TIMES trademark since 1922 and the LATIMES.COM trademark since 1996. *The Los Angeles Times* enjoys both national esteem and noteworthy commercial success, ranking among the largest newspapers in the United States by print circulation and routinely winning awards for its coverage of late-breaking events, including the Pulitzer Prize for Breaking News Reporting in 2023.

53. McClatchy is the owner of several valid, protectable marks including THE MIAMI HERALD, THE IDAHO STATESMAN, STAR-TELEGRAM, and FORTH WORTH STAR-TELEGRAM. McClatchy and its predecessors-in-interest have used the trademark THE MIAMI HERALD in connection with newspapers since December 1910, the trademark THE IDAHO STATESMAN since June 1969, and STAR-TELEGRAM-inclusive marks since December 1909. *The Miami Herald* is known for the credibility of its factual reporting, winning 24 Pulitzer Prizes and several Online Journalism Awards for its breaking news and explanatory reporting. Over its 120-year history, the *Fort-Worth Star Telegram* has developed a reputation for incisive reporting

and innovation, and it is credited as the nation's oldest continuously operating online newspaper. The *Idaho Statesman* has long been associated with excellent breaking news coverage of issues affecting Idahoans, and in 2008 was named a finalist for the Pulitzer Prize in Breaking News Reporting for its coverage of the arrest of the U.S. Senator from Idaho. Each of these marks is associated with top-tier journalism and, because of that track record, represents significant public trust.

54. Newsday is the owner of the valid, protectable trademark NEWSDAY, which it has used in connection with news articles since September 3, 1940, when the first edition of *Newsday* was published in Hempstead, New York. Through its reliable reporting on news affecting the residents of Nassau and Suffolk County and its prompt coverage of how national and international events affect Long Island, Newsday has steadily gained public trust. While Newsday's focus is local, it has received national recognition for the quality of its journalism, earning three Pulitzer Prizes for Breaking News Reporting and two gold medals for Public Service.

55. Torstar is the owner of the valid, protectable TORONTO STAR trademark, which it has used continuously in connection with the sale of newspapers and news articles since 1971. The *Toronto Star* is the largest Canadian newspaper by circulation and, through its popular website and mobile applications, reaches a global audience, including U.S. subscribers. After over fifty years of use and promotion, Torstar's customers in the United States understand the *Toronto Star* as a dependable source of news on international affairs—and recognize the TORONTO STAR mark as uniquely identifying news articles that originate with the *Toronto Star*.

56. Vox Media is the owner of several valid, protectable marks, including VOX, which it has used without interruption to sell and promote online news and opinion articles since the launch of *Vox* on April 5, 2014. In just over a decade, Vox Media has put the *Vox* mark on

explanatory journalism. Vox Media has won five SEAL (Sustainability, Environmental Achievement & Leadership) Awards for its environmental journalism and, in 2023, Vox Media journalists were named as finalists for three Online News Association awards, including an award honoring Excellence in Feature Writing. The consistent quality and considerable success of the *Vox* brand has led consumers nationwide to associate the VOX trademark with in-depth, insightful reporting on current events.

57. In addition, in 2019, Vox Media acquired as a wholly owned subsidiary New York Media, LLC and the entire interest and goodwill in the trademarks NEW YORK, GRUB STREET, and VULTURE. Vox Media and its predecessor-in-interest have continuously used the mark NEW YORK in connection with news articles since the first issue of *New York* Magazine was published on April 8, 1968; the mark GRUB STREET in connection with online articles on topics of culinary interest since 2006; and the mark VULTURE in connection with online articles covering culture and entertainment since 2007. In the decades since *New York*'s founding, it has won 53 National Magazine Awards from the American Society of Magazine Editors (ASME). *Vulture* and *New York* have produced two recipients of the Pulitzer Prize for Criticism. Similarly, *Grub Street* is an acclaimed source of food and restaurant-related journalism, earning the James Beard Award for Group Food Blog in 2011 and 2015.

58. In addition to their common-law rights in their respective marks, accrued over years of use and promotion, Publishers Advance, Condé Nast, The Atlantic, Axel Springer Companies, Forbes, The Guardian, LA Times, McClatchy, Newsday, and Vox Media own valid and subsisting federal trademark registrations, as identified in Exhibit C.

59. As a result of Publishers' continuous, exclusive, and longstanding use of these trademarks in connection with their publications, consumers have come to recognize Publishers'

trademarks as exclusively identifying Publishers’ goods and services. Moreover, these brands are known to the public as high-quality sources of reliable and informative content, engendering incalculable goodwill. The reputations of these brands are essential to Publishers’ ability to deliver news and media pieces that the public will want to consume and pay for.

II. Cohere and Its Business

60. Founded in 2019, Cohere is in the business of developing, operating, and licensing AI models. As of July 2024, Cohere was valued at \$5.5 billion. Its lead funders include other large corporate technology companies such as Oracle, NVIDIA, and Salesforce.

61. Cohere’s primary product is its suite of LLMs referred to as the Command Family of models—including Command, Command R, and Command R+. These LLMs are trained on vast amounts of text and as a result can generate text-based, natural language responses to user queries. These models can thus be used, as Cohere describes, for “powering conversational agents, summarization, copywriting, and similar use cases.”

62. Cohere trains and hosts all its models in the United States. Cohere’s services and data are hosted on servers located in the United States, which it has confirmed in its technical documentation available to public.

63. One way to access the Command LLMs is through Cohere’s online Chat interface, first made available on Cohere’s website in July 2023 under the name Coral. Users can chat with Command models for free on Cohere’s website or via an API intended for use with software. Cohere makes free trial access available to serve its goal of capturing more paying customers.

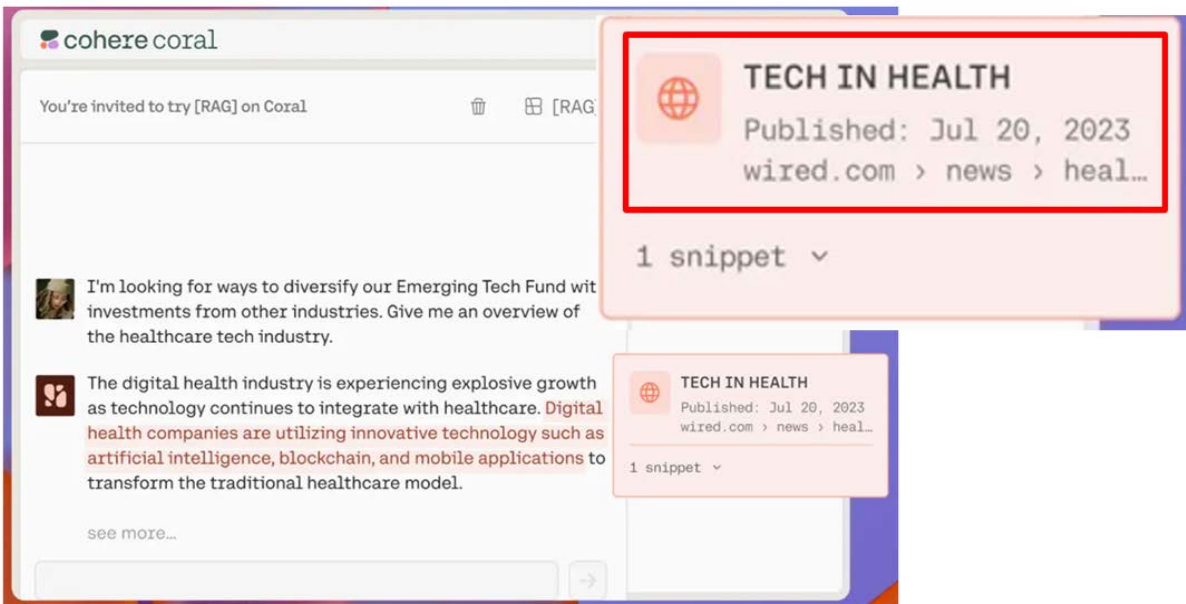
64. Cohere offers various paid options for customers to use instances of the Command Family of models that Cohere itself operates. Customers can access the Chat interface, paying according to the volume of input and output—that is, according to the length of prompts submitted

and output received. Cohere also offers customers the ability to run their own instances of the Command Family of models in computing infrastructure separately arranged by the customer. In this case, Cohere provides models under negotiated agreements.

A. Cohere Emphasizes That It Provides the “Latest News”

65. Cohere tries to differentiate its product in the AI marketplace as especially suited to the business community. Cohere markets Command (and its Chat function) as a “knowledge assistant” and focuses on techniques to reduce AI’s tendency to hallucinate or give incorrect answers. Command is designed to shortcut research and content analysis, whether the object is a customer’s own documents or copyright-protected journalism.

66. Cohere has specifically called out and emphasized Command’s news reporting function. For example, as shown below, when Cohere introduced Chat (formerly Coral) in July 2023, the very first example it marketed of its “knowledge assistant” at work shows a response generated by reliance on a piece from *Wired*, a Condé Nast publication.²⁹

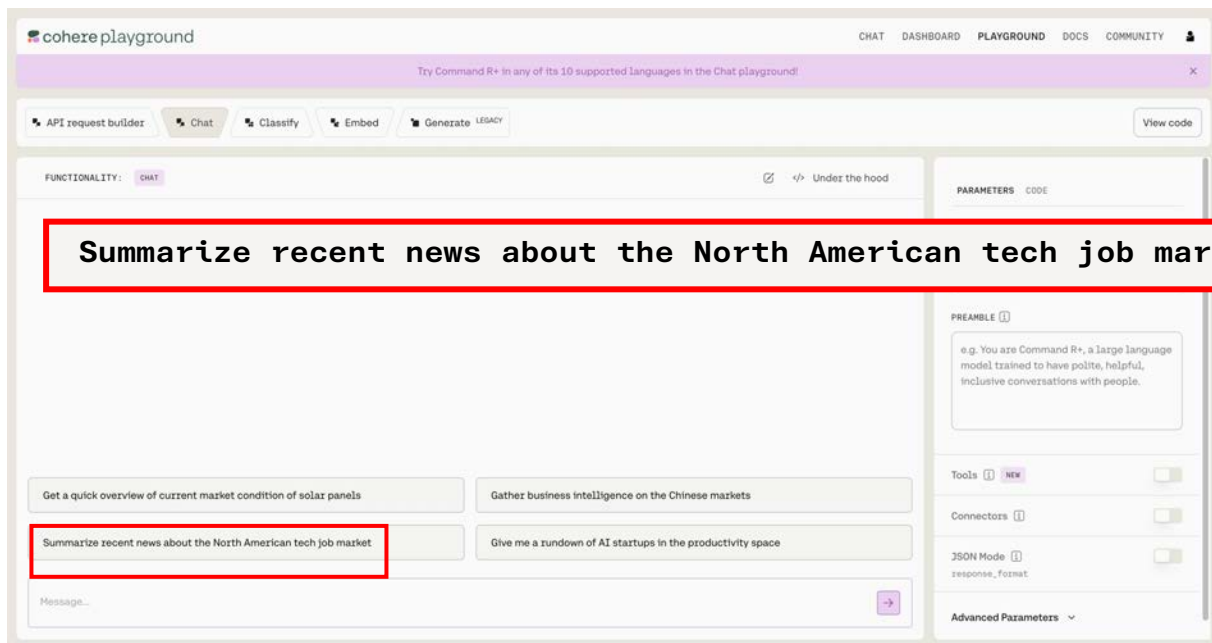


²⁹ *Introducing Coral, the Knowledge Assistant for Enterprises*, COHERE (July 25, 2023), <https://cohere.com/blog/introducing-coral>.

67. In a marketing pitch a few months later, in September 2023, Cohere again made the case for Chat as a knowledge assistant: “[A] developer building a market research assistant can equip their chatbot with a web search to access the latest news about trends and competitors in their space.”³⁰

68. Cohere has continued to solicit customers by promoting its Command models as tools to access news. For instance, in September 2024, Cohere announced on its website marketing that the Cohere AI app was available in Slack and touted its ability to “keep you up to date with the latest news.”³¹

69. On its Chat interface, Cohere suggests to consumers that its models provide summaries of the latest news articles. In the Playground tab of Cohere’s product demo for its Command models, Cohere even prepopulates the interface with a request to summarize recent technology news, inviting prospective customers to use the models to access news stories.



³⁰ *Introducing Chat with Retrieval-Augmented Generation (RAG)*, COHERE (Sept. 28, 2023), <https://cohere.com/blog/chat-with-rag>.

³¹ *The Cohere AI App is available in Slack*, COHERE (Sept. 16, 2024), <https://cohere.com/blog/slack-cohere-ai-app>.

70. But Command is incapable of performing its own original research. It invests no resources into news gathering in the field and no has writers, fact-checkers, or editors on staff. And while Cohere actively courts customers with promises to treat their proprietary content safely, it has not bothered to partner with media publishers to fairly acquire the valuable and reliable journalism that powers the Command models' responses. Publishers have not permitted or otherwise authorized Cohere to use their respective copyrighted works, their trademarks, or otherwise engage in any of the activity complained of herein.

B. Cohere's Training of Command and Cohere's RAG Activity

71. Generative AI LLMs are built to respond to user prompts and queries in a seemingly human manner by following patterns derived from the text on which they were trained. Larger and more capable models require training on larger amounts of text. Accordingly, Cohere copies massive amounts of text from the internet, including Publishers' works, to create a dataset on which to train its models. Cohere collects this material by copying and downloading, also known as "scraping," that text directly from websites and onto its servers, using automated tools, such as web crawlers and other bots. Cohere also prepares its training dataset by working from collections prepared by third parties, which in turn have been harvested through web scraping. This vast collection of text is the basis of a "corpus" that Cohere then uses to train its Command model.

72. After harvesting the training corpus, Cohere "cleans" the text to remove unwanted material such as duplicates, offensive language, or any other material inconsistent with its business purpose. Next, Cohere converts the text into "tokens." Tokens are words or parts of words or punctuation. Tokens enable the efficient processing of language in a way the model can understand and analyze. Then, Cohere trains the model to deliver outputs in response to queries. The model is "finetuned" by Cohere using both human and AI feedback, to calibrate the model for specific tasks

or use cases, which may also use or depend upon unauthorized copies. Cohere also includes certain “safeguards” at this stage, which attempt to prohibit the model from delivering offensive, dangerous, or otherwise undesirable material. In addition, Cohere may apply filters and guardrails external to the model, both on user prompts and model outputs, to limit various sorts of activity. Cohere makes multiple additional copies of the text in the training corpus across these steps in the development process.

73. Once trained, Cohere’s models generate natural language outputs consistent with the text on which it was trained. But these models only have information up to the time Cohere harvested the training content. For this reason, AI models get stale very quickly.

74. In response to some prompts, LLMs can and will perfectly regurgitate training material. In other cases, LLMs fail to accurately represent the content on which they were trained. The result is LLMs are not guaranteed to provide accurate information. Worse, LLMs will “hallucinate,” or fabricate, plausible sounding but inaccurate information in response to user prompts.

75. To reduce hallucinations and provide more useful and timely responses to user queries, Cohere’s approach is to rely heavily on “retrieval augmented generation” (“RAG”), a term coined by Dr. Patrick Lewis, now a researcher at Cohere. RAG supplements the user’s prompts with additional information from external data sources so that the Command model has material from which to generate a response. Cohere provides a “Web Search connector” that it hosts and manages so that it searches for and selects web pages to be used to supplement its response to user prompts. That is, Cohere will use third-party websites, including the Publishers’ sites, as content sources for RAG.

76. Cohere relies heavily on trusted journalism sources to shore up the authority of its responses. As Cohere’s CEO Aidan Gomez explained in a letter to employees and shareholders, Cohere believes that a “key differentiator” for its models is the ability to receive “verifiable answers.”³² Publishers’ scrupulously researched and fact-checked stories are high-quality sources for Cohere’s models. Indeed, Command’s RAG feature routinely returns copies of Publishers’ copyrighted content as part of its responses to user queries, including same-day, breaking news pieces and verbatim copies.

77. Marrying the conversational ability of LLMs with the knowledge base of the internet, or other databases, makes a more compelling product than an LLM alone. Indeed, the ability to pull content from external sources for use in processing by Command models is essential to Cohere’s business and its core differentiating feature in the generative AI marketplace. Generative AI models that rely exclusively on training data to generate responses are unsatisfactory to users who require up-to-the-minute, accurate information, especially business customers. Indeed, one of Cohere’s founders has stated: Hallucinations from models without RAG are “actually a feature, not a bug.”³³

C. Interacting with Command via Cohere’s Chat

78. Use of Cohere’s services requires a customer to agree to several “Safety” and “Legal” terms and advise Cohere if any “sensitive use cases” are intended. Cohere’s written usage policy purports to prohibit Cohere’s services from being used for, among other things, plagiarism, deceit, and misrepresentation.

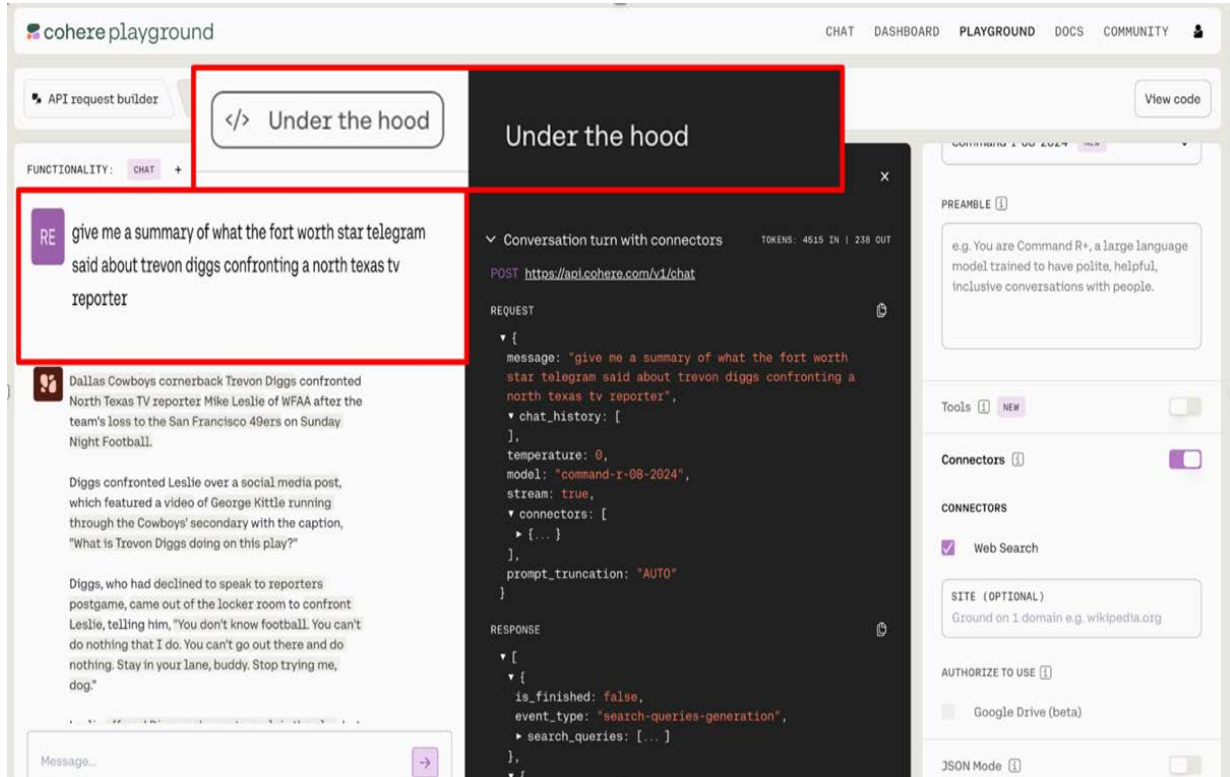
³² Aidan Gomez, *Where Enterprise AI is Headed*, LINKEDIN (Dec. 5, 2024), <https://www.linkedin.com/pulse/where-enterprise-ai-headed-aidan-gomez-dznjf>.

³³ *Cohere’s Ivan Zhang on Foundation Models, RAG, and Feedback Loops*, MADRONA (Aug. 22, 2023), <https://www.madrona.com/cohere-ivan-zhang>.

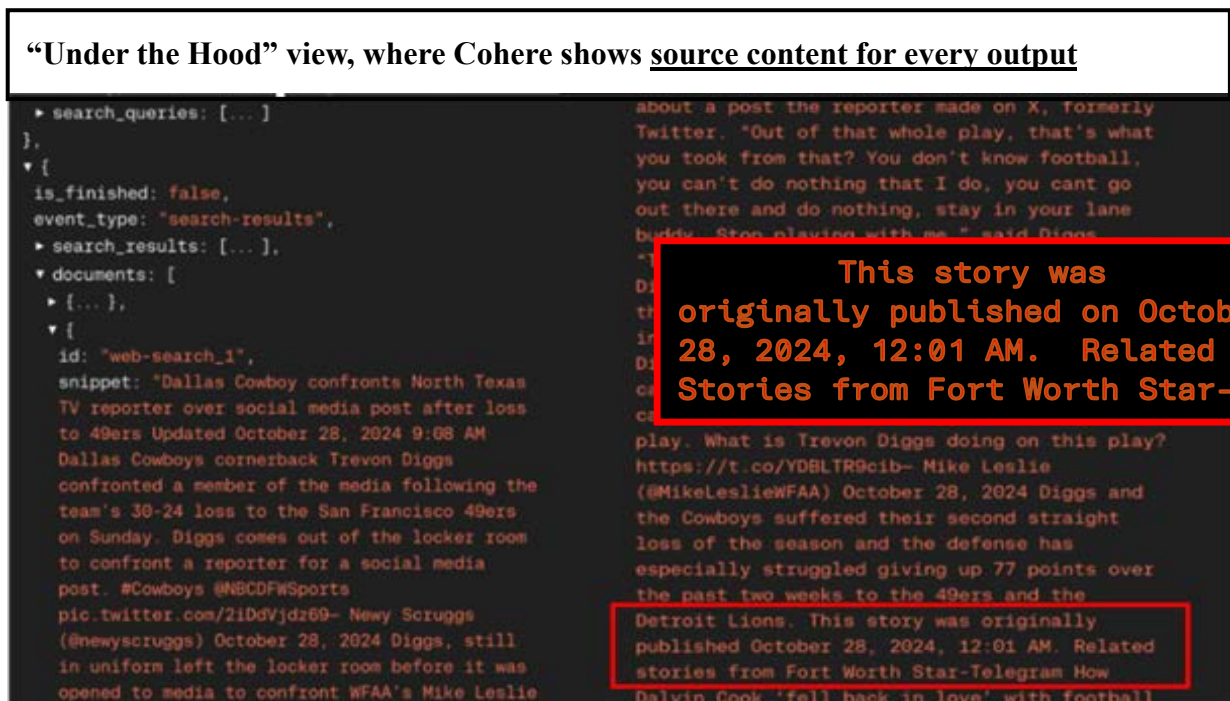
79. Users can interact with Command via Cohere’s Chat service in a variety of ways. A user must first create an account with Cohere by signing up with an email and password. The user then can immediately begin interacting with the Chat service at <https://coral.cohere.com>, where two different interfaces are available: a basic one for ordinary users and a “Playground” for more technical users. In the basic interface, Cohere offers two options: “Just Chat” or “Use Tools.” Just Chat will allow the user to input a query, and the model will deliver an output based solely on what it “learned” from its training corpus. Use Tools will give the user the option for the model to incorporate content found via a web search in forming its response. Cohere does not use the term RAG explicitly in this basic interface, but by selecting the web search option, the user will in fact be turning on the RAG feature.

80. The “Playground” interface provides the Chat dialog box, but with more extensive options including a feature called Under the Hood. This option allows the user to see not only the specific source(s) from which Cohere copied to answer the query (which is also provided in the basic interface with the Use Tools option), but also the underlying text of those sources. The user can expand Under the Hood to view the exact underlying documents on which Cohere relied to generate the response. Cohere refers to these sources as “snippets,” but to be clear—these “snippets” are generally the full text of every source on which the output was based.

81. Below is a screenshot of the Cohere Playground showing a 2024 request for the summary of an article from the *Star-Telegram*, the beginning of the model’s response in the output section, and the details of how the result was generated displayed in the Under the Hood panel:



The Under the Hood section can be expanded so the user can read a copy of the full underlying document(s) Cohere obtained through its web search feature and used in forming a response:



III. Cohere's Infringement of Publishers' Copyrighted Works

82. Cohere copied Publishers' works to train and implement its Command Family of models and services built on them. Cohere also delivered outputs that are either full verbatim copies, substantial excerpts, or substitutive summaries of Publishers' works, including to Publishers' investigators, for each of the Works in Suit on Exhibit A. Each of these steps constitute separate acts of infringement and Cohere continues to engage in that infringement.

A. Copying During Training

83. Cohere engages in wholesale, verbatim copying of Publishers' works to train its models. Cohere CEO Aidan Gomez has stated that Cohere scrapes the web and feeds billions of eBooks and webpages including WordPress, Tumblr, Stack Exchange, Genius, the BBC, Yahoo!, and the New York Times to the models to train them.³⁴ Cohere also uses C4, a version of Common Crawl data, to train its models.

84. The Common Crawl Foundation is an organization that crawls the internet, extracts large portions approximately every month, and provides that content to the public at no cost. Google researchers defined a methodology for creating a subset useful for AI training, known as the Colossal Clean Crawled Corpus ("C4"), from a Common Crawl dataset. This content, based on data which Common Crawl provides for free, is a very attractive dataset for AI developers. Unfortunately, Common Crawl does not differentiate between copyrighted and public domain material.

85. Publishers' works are among the most highly represented sources in an iteration of C4 dataset prepared and distributed by the Allen Institute for AI. The websites for the *Los Angeles Times*, *The Guardian*, and *Forbes* ranked fifth, sixth, and eighth, respectively, of the most

³⁴ Kyle Wiggers, *OpenAI rival Cohere launches language model API*, VENTUREBEAT (Nov. 15, 2021), <https://venturebeat.com/uncategorized/openai-rival-cohere-launches-language-model-api>.

represented domains in C4, with *Business Insider* ranked 16th and *The Atlantic* ranked 19th. The News and Media category ranks third across all categories of content. Publishers, particularly their magazine properties, are also well represented in categories like “Business & Industrial,” “Arts & Entertainment” and “Hobbies & Leisure.” Publishers, particularly their magazine properties, are also well represented in categories like “Business & Industrial,” “Arts & Entertainment,” and “Hobbies & Leisure.” Many of the Works in Suit are in C4.

86. In addition to C4, Cohere crawls particular sites to copy additional content, including at least some of Publishers’ websites. This is no accident. Cohere chooses Publishers’ articles because they are a rich source of trustworthy, expressive, and newsworthy content, highly valuable to an LLM’s dataset.

87. Cohere’s use of Publishers’ works in training may be confirmed through the sources referenced Under the Hood in Command outputs. Dates shown on the source material are consistent with earlier downloading for use in training, rather than downloading in connection with a current request.

88. Copyright is not an “opt out” system. Cohere has an obligation not to use Publishers’ copyrighted content without authorization regardless of whether Publishers have taken affirmative steps to block Cohere’s crawlers. And while Cohere claims it pays “a lot” for content in some cases,³⁵ it has not paid Publishers for training data or otherwise obtained their permission.

89. Cohere claims to dedicate much of its time to properly curate content to avoid biases, misuse, and harmful text. But it has not filtered out unlicensed copyrighted content from

³⁵ Nilay Patel, *AI will make money sooner than you’d think, says Cohere CEO Aidan Gomez*, THE VERGE (June 10, 2024, 10:30 AM EDT), <https://www.theverge.com/24173858/ai-cohere-aidan-gomez-money-revenue-llm-transformers-enterprise-stochastic-parrot>.

its datasets. On the contrary, the datasets on which Cohere trained its models and on which its Command models rely to generate text include Publishers' works.

90. Cohere knows that it engaged in massive infringement of Publishers' copyrighted works to train Command. Cohere has complete control over the process for training, fine-tuning, and testing Command. It chose C4 as a component of its dataset and supplemented that data with its own crawling, knowing that it was extracting (that is, copying) copyrighted material. Publishers have put Cohere on notice that it was not authorized to use their works by including copyright notices with their works and terms of service on their websites. Cohere has even proceeded notwithstanding directives in robots.txt files and receiving a cease-and-desist letter. In sum, Cohere can exclude Publishers' copyrighted works from inclusion in its training but has chosen not to.

B. Reproduction, Distribution and Display of Works Used in Training

91. As discussed in the next section, with the web search connector active, Cohere delivers copies of Publishers' works it obtains, including recent articles that post-date the corpus Cohere used to train Command. Cohere provides copies of Publishers' work in response to a variety of common-sense, natural-language user queries, such as "give me a copy of X article" or "tell me about Y." Sometimes, however, Cohere will deliver a copy of the work from an older copy. While discovery from Cohere will shed more light, this may reflect that Cohere is providing to its users a copy of a copy that it retained from its training corpus.

92. Below is a 2024 request for a January 2019 *Newsday* article that is also in the C4 dataset. Cohere copied the article verbatim into its Chat output, and the metadata displayed in the Under the Hood panel indicates Cohere used a December 2021 copy of the authentic article.

PROMPT TO COHERE:

Give me the article PSEG LI: We don't have to pass along federal tax savings published by Newsday.



Emphasis added in blue font marks Cohere's verbatim copying from the Newsday article

COHERE'S CHAT OUTPUT:³⁶

PSEG Long Island's unique contract with the Long Island Power Authority exempts the company from a state order to share its windfall from federal tax reductions PSEG and state officials said. LIPA in its annual budget reported it received a total of \$6 million in savings from contractors who saw significant reductions from President Donald Trump's tax reform bill which lowered the corporate tax rate to 21 percent from 35 percent. The Public Service Enterprise Group of Newark PSEG's parent reported in financial filings last year that it expected to record a one-time noncash earnings benefit in the range of \$660 million to \$850 million as a result of the tax cuts. Individual PSEG ratepayers in New Jersey will see an average benefit from the utility's \$114 million tax break of ...

Under the Hood reflects:

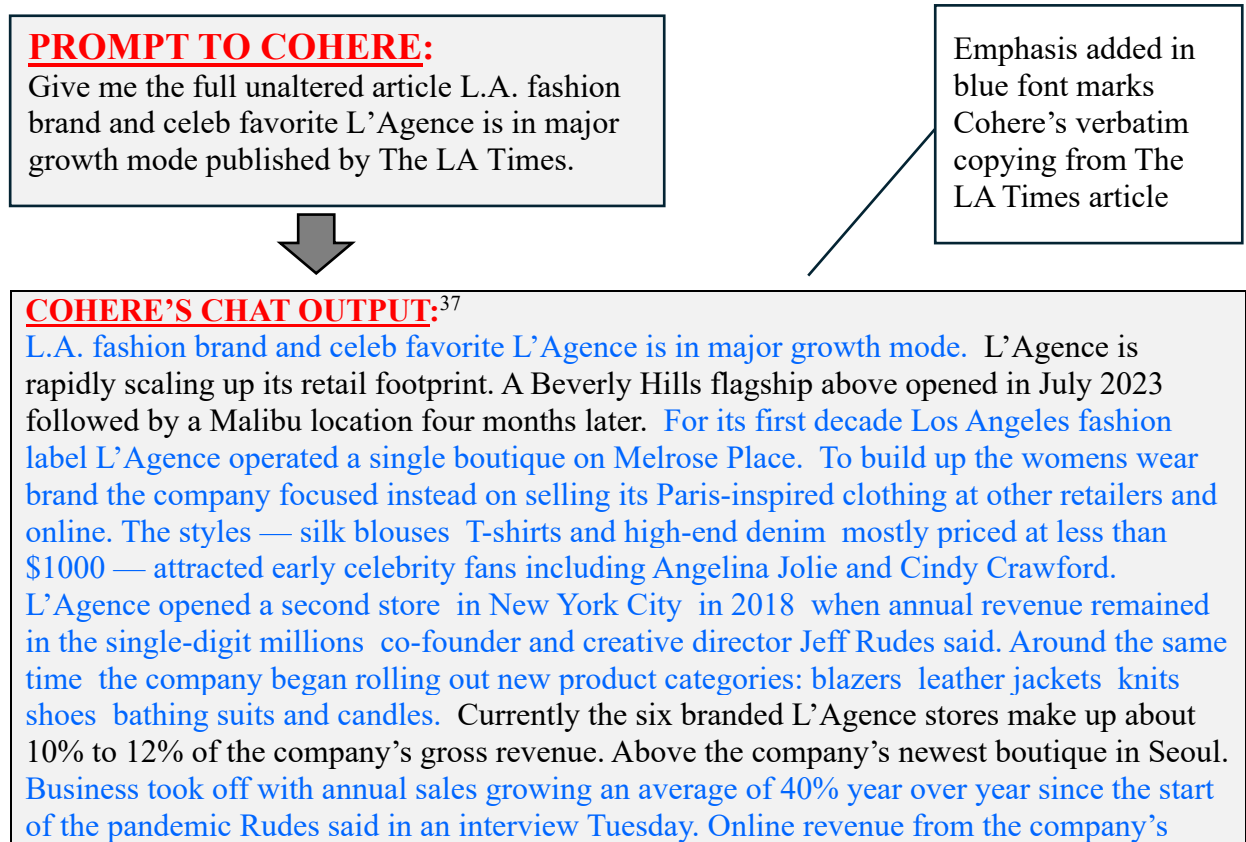
- Document Timestamp: 2021-12-10T00:33:49
- Document "Snippet" Further Shows Article Copied on 12/10/2021: Includes text, "Skip to content Menu Today's paper 12/9/2021 Good Evening 12/9/2021 Good Evening"
- Source pulled originally from a Newsday URL that is no longer active: <https://www.newsday.com/long-island/pseg-long-island-trump-tax-cut-1.26104276>

93. When Cohere delivers a copy of an article to a user as part of a Chat output—whether a verbatim copy, excerpt, or close paraphrase—Cohere makes a full copy of the source article before incorporating it into its response. Thus, Cohere makes yet another unauthorized reproduction of the piece, apart from copies made during training. Further, Cohere makes an unauthorized reproduction, distribution, and display when Cohere delivers that copy to the user. Cohere knows that its Command-model-based Chat service generates and delivers copies of Publishers' works to end users, because Cohere, contrary to its own policies against infringement, programmed its system to do exactly that.

³⁶ Cohere's infringing output continues beyond the abbreviated portion copied into the body of this Complaint. For a full copy, see Exhibit B, Example 40.

C. Reproduction, Distribution, and Display of Newer Works

94. As mentioned above, when the RAG feature is turned on, Cohere consistently delivers copies of Publishers’ copyrighted works either within the body of its response or in conjunction with the response (as shown in the “Under the Hood” panel). It does this whether the user asked explicitly for the specific work or not. As noted, Cohere sometimes bases its response on copies with timestamps indicating they were retrieved months or years in the past. However, copies can include newer articles that post-date Command’s training dataset, including pieces published on the day of Cohere’s retrieval and response through the Chat interface. For example, when asked for a particular October 2024 article in the *Los Angeles Times*, Cohere delivered the verbatim with only slight word and punctuation changes.



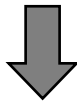
³⁷ Cohere’s infringing output continues beyond the abbreviated portion copied into the body of this Complaint. For a full copy, see Exhibit B, Example 30.

website alone last year totaled “well over \$100 million.” Now L’Agence is rapidly scaling up its retail footprint. A Beverly Hills flagship opened in July 2023 followed by a Malibu location four months later. This summer the company turned its original ...

95. Cohere delivers verbatim texts of Publishers’ copyrighted articles even when asked generally for information about a particular topic rather than a specific piece. For example, in response to a 2024 prompt asking generally about financial strains on the public transit system in Miami-Dade County, Cohere delivered a complete article from the *Miami Herald*, with only a minor change to one sentence:

PROMPT TO COHERE:

Tell me about the budget strains on future transit funding in Miami-Dade county.



Emphasis added in blue font to mark Cohere’s verbatim copying from the *Miami Herald*

COHERE’S CHAT OUTPUT:³⁸

Miami-Dade County is saving each year to expand transit, but the 2025 budget proposal by Mayor Daniella Levine Cava freezes payments to those reserves for a year. The 2025 budget takes \$16 million away from future rail projects to fill gaps in a spending plan that faces its first vote on Thursday evening. For the first time since the transit savings program began six years ago, Miami-Dade would not fund the required payment in 2025 for the taxing district created in 2018 around the county’s Metrorail and Metromover tracks. There is \$82 million in the fund now, and the Levine Cava budget blocks the ...

96. Cohere’s infringing copies directly supplant Publishers’ ability to benefit commercially from the transaction, such as through individual or enterprise-level subscription fees that Publishers would charge to access their articles or through advertising revenue Publishers

³⁸ Cohere’s infringing output continues beyond the abbreviated portion copied into the body of this Complaint. For a full copy, see Exhibit B, Example 36.

would earn from web traffic to their articles. In fact, Cohere delivers copies of articles even when they are protected by paywalls on the Publishers' websites.

97. In addition, in any instance where Cohere uses its web search RAG function to obtain a work from Publishers' websites, it must make a copy of the piece before incorporating it into its response. Thus, Cohere makes an unauthorized reproduction of the piece, in addition to its other infringing activities in delivering outputs of newer works.

98. Cohere's infringement of Publishers' works is knowing and intentional. As discussed above, the outputs delivered by the Command model-based Chat service are the product of Cohere's choices and conduct. Cohere selected and gathered the content to use in training its Command models, controls the web search connector that provides content for use with RAG (including whether that feature ignores paywalls and robots.txt directives), and controls the filters and guardrails it applies—or fails to apply—to its Chat service.

99. Indeed, Cohere even delivers verbatim texts of Publishers' copyrighted articles close in time to when Publishers first publish them. The example below shows a request for a *Business Insider* article that was published on the Publisher's website at 11:08 am EDT, a reasonable search query for a reader seeking breaking news. By 1:14 pm EDT, Command delivered almost the entire article when prompted, apart from small word and punctuation changes. Verbatim copying is shown in blue. The information in the Under the Hood panel shows that Cohere used a copy of the article from the Publisher's website made at 12:55 pm EDT, less than two hours after publication and prior to the 1:14 pm EDT prompt.

PROMPT TO COHERE:
Give me the article “Disney’s CEO plan sparks speculation on what the drawn-out Bob Iger succession means” published by *Business Insider*.

Emphasis added in blue font to mark Cohere’s verbatim copying from *Business Insider*



COHERE’S CHAT OUTPUT:³⁹
Disney's CEO plan sparks speculation on what the drawn-out Bob Iger succession means
Lucia Moses 2024-10-22T15:08:02
Bob Iger is set to leave the company at the end of 2026. Slaven Vlasic/Getty Images
This story is available exclusively to Business Insider subscribers. Become an Insider and start reading now. Have an account? Log in.
Disney said Monday it would name a new CEO in early 2026 and named James Gorman as board chairman. Bob Iger's previous succession plan failed, leading to his return as CEO in 2022. Some close to Disney are wondering how his next exit will be handled [. . .extraneous content removed . . .]
News about succession plans more than a year in the future which fail to name any potential successors wouldn't merit a press release at many other companies. But Disney isn't an ordinary company, considering its high public profile and Bob Iger's history of extending his tenure. Disney's announcement Monday that it would name its next CEO in early 2026 and its appointment of a new board chairman, former Morgan Stanley CEO James Gorman has drawn a mix of reactions and questions about the future. ...

2024-10-22T16:55:53 (12:55 pm EDT, assuming Cohere uses UTC)

100. Tellingly, the output also states “This story is available exclusively to Business Insider subscribers. Become an Insider and start reading now,” all the while providing the full article to any user who asks for it, whether they have a *Business Insider* subscription or not.

D. Unauthorized Distribution and Display of Substitutive Summaries

101. While Cohere often responds to a request for a specific article with a complete, verbatim copy or substantial portions thereof, sometimes Cohere instead delivers a substitutive

³⁹ Cohere’s output continues beyond the abbreviated portion copied into the body of this Complaint. For a full copy, see Exhibit D, Example 10.

summary of the requested work, including when asked to provide a summary about a topic as depicted by a particular publisher, or when asked about a topic more generally.

102. Cohere first copies the article in full to generate the competing summary, displaying the copy in the Under the Hood panel. Then, Cohere copies from and summarizes that full copy article, often usurping so much of the relevant article that the user has no need to consult the original work. In the example below, Cohere was asked “Tell me about portland public schools not allowing teachers to talk about political matters.” Cohere first copied a September 2024 piece from Publisher Advance’s newspaper *The Oregonian/OregonLive*, “Portland Public Schools quietly adopts policy barring teachers from ‘political or personal’ classroom displays.”

103. Here is the text of the full article as Cohere presented it in the Under the Hood panel:

```

{
  "id": "web-search_0",
  "snippet": "Portland Public Schools quietly adopts policy barring teachers from 'political or personal' classroom displays\n\nUpdated: Sep. 12, 2024, 7:02 a.m.\n\nPublished: Sep. 12, 2024, 7:00 a.m.\n\nA Black Lives Matter flag flies in front of Grant High School in Northeast Portland. The school was among the first testing grounds of a new policy that restricts classroom posters and art to material that is connected to curriculum. Photo by Beth Conyers/Portland Public Schools\n\nJulia Silverman | The Oregonian/OregonLive\n\nA new policy governing what Portland Public Schools educators are – and are not – allowed to put on the walls of their classrooms is drawing pushback from its vocal teachers' union.\n\nUnder the policy, items on display in classroom walls and bulletin boards must be related to approved curriculum or district sponsored events. The rules specify that classroom spaces "cannot be used for an employee's personal expression, whether that is related to a political or personal issue.\n\nThe new guidance was adopted without fanfare in August as an administrative directive, meaning that it did not require approval from the school board or a public airing.\n\nThe change comes after districtwide upheaval last spring tied to the Israeli-Hamas conflict, after a group affiliated with the Portland Association of Teachers published a controversial guide to teaching and organizing in support of Palestinian statehood. Suggested lessons included encouraging students to pray to Allah on behalf of Palestinians and writing letters to the U.S. president to urge an end to funds for Israel.\n\nNone of the first tests of the new policy came when administrators at Grant High confirmed the removal of "Stop the Genocide" posters from a social studies teacher's doorway in early September. The poster's message referred to the deaths of tens of thousands of Palestinians killed in Israeli military actions after the Hamas hostage takings last October.\n\nIn an email about the situation sent to a concerned parent and shared with the Oregonian/OregonLive, Grant High principal James McGee said the posters were taken down because they were "in violation of PPS policy regarding political speech.\n\nThe teacher, Jeremy Reinholt, declined to comment on the situation. Among the classes he is assigned to teach this year is modern world history.\n\nPortland Association of Teachers President Angela Bonilla said via written statement that she believes the policy "violates our collective bargaining agreement. In addition, it is unworkable, overly broad and vague.\n\nThe union filed a grievance with the district over the new policy in August, contending that it conflicts with contractual language on academic freedom, which allows teachers to introduce controversial topics that are relevant to their courses, and PAT's right to post union related materials in schools.\n\nPAT has asked its members to regularly take pictures of walls in their classrooms and in the hallways and cafeterias and on doors, to document how the new policy is being enforced. According to the rules, all posted material in such common spaces must be approved by an administrator and can only be displayed in certain areas.\n\nDistrict officials said through a spokesperson that the policy change has been in the works for over a year, in response to concerns raised by schools about murals in their buildings, some of which date to the Works Progress Administration era. Some have been removed in recent years, including murals at Grant that depicted idealistically harmonious interactions between Indigenous peoples and white settlers.\n\nOur policy is focused on ensuring the classroom spaces continue to be used for curriculum and PPS-related content," the district's statement said, in response to a question about the impetus for the new policy.\n\nThe administrative directive is content neutral, and we believe supports the goal of remaining focused on creating a rich educational environment for students.\n\nTwo years ago, a judge in Yamhill County voided a Newberg-Dundee School District policy that banned employees from displaying "controversial" or "political" signs or symbols after the ACLU filed suit on behalf of a district employee. Judge Cynthia Easterday said then that Newberg's policy was in violation of Oregon's broad constitutional protections around freedom of expression.\n\nBoard members in the Newberg-Dundee district, which is about an hour outside of Portland, had originally sought to ban teachers from displaying Black Lives Matter and LGBTQ+ pride flags, though control of the district's school board has since flipped to more progressive-leaning members.\n\nSchool leaders in Portland said their policy wouldn't infringe on a teacher's ability to display similar material.\n\nThe rainbow flag and BLM poster are district-approved symbols of inclusion to often-marginalized students," district spokesperson Valerie Feder said. "Posters advocating for specific positions on political positions are not student-centered in that they are not rooted in our educational mission or curriculum. Personal expression by employees is not in furtherance of PPS's academic purposes.\n\nBut Bonilla said she thought the policy left questions unanswered.\n\nFor instance, if an LGBTQ educator shares a photo of their family with a poster that reads, 'Families come in all shapes and sizes,' would that be inappropriate?" she asked.\n\nGrant High parent Shoshana Kedem said she supported the policy's intent.\n\nIf someone complains that it makes it hard for them to be in a learning environment, then those things should come down," she said of the Stop the Genocide posters at her child's school. "If there were Israeli flags and signs about supporting the occupation, I think people would say the same thing.\n\nJulia Silverman covers education for The Oregonian/OregonLive.com. Reach her via email at jsilverman@oregonian.com. Follow her on X.com at @jrsilverman.\n\nIf you purchase a product or register for an account through a link on our site, we may receive compensation. By using this site, you consent to our User Agreement and agree that your clicks, interactions, and personal information may be collected, recorded, and/or stored by us and social media and other third-party partners in accordance with our Privacy Policy.\n\nUse of and/or registration on any portion of this site constitutes acceptance of our User Agreement, (updated 8/1/2024) and acknowledgement of our Privacy Policy, and your Privacy Choices and Rights (updated 7/1/2024).\n\n© 2024 Advance Local Media LLC. All rights reserved (About Us). The material on this site may not be reproduced, distributed, transmitted, cached or otherwise used, except with the prior written permission of Advance Local.\n\nCommunity Rules apply to all content you upload or otherwise submit to this site.\n\nYouTube's privacy policy is available here and YouTube's terms of service is available here.\n\nOpt out of the sale or sharing of personal information\n\nIf you opt out, we won't sell or share your personal information to inform the ads you see. You may still see interest-based ads if your information is sold or shared by other companies or was sold or shared previously.",
  "timestamp": "2024-09-12T17:12:57",
  "title": "Portland Public Schools quietly adopts policy barring teachers from 'political or personal' classroom displays - oregonlive.com",
  "url": "https://www.oregonlive.com/education/2024/09/portland-public-schools-quietly-adopts-policy-barring-teachers-from-political-or-personal-classroom-displays.html?outputType=amp"
}

```

104. Cohere summarized the above article as follows (green font marks paraphrasing from *The Oregonian*):

COHERE’S CHAT OUTPUT:⁴⁰
 In August 2024, Portland Public Schools (PPS) adopted a new policy that restricts teachers from displaying "personal or political" messages on the walls of their classrooms. The policy states that displays must be tied to approved curriculum or district-approved events. The change in policy came after a dust-up in the district in June 2024 regarding the Israeli-Palestinian conflict. The Portland Association of Teachers (PAT) published a controversial guide in support of Palestine, which included heavily pro-Palestinian suggestions. This outraged some members of the local Jewish community. The new policy has faced backlash from the teachers' union, which filed a grievance against the directive. The union argues that the policy violates the union's collective bargaining agreement with the district and appears to violate the safeguards on academic freedom in its contract. PAT President Angela Bonilla stated that the policy "violates our collective bargaining agreement. In addition, it is unworkable, overly broad and vague." ...

105. This output heavily paraphrases from the text of the original article. As detailed below and shown in green text above, Cohere extracted from *The Oregonian* the specifics of a new rule in Portland regarding the display of political messages, the genesis of the new rule, and the reaction of the teachers’ union including the details of a grievance action:

<u>COHERE’S CHAT OUTPUT</u>	<u>ORIGINAL ARTICLE</u>
Portland Public Schools (PPS) adopted a new policy that restricts teachers from displaying "personal or political" messages on the walls of their classrooms.	A new rule governing what Portland Public Schools educators are — and are not — allowed to put on the walls of their classrooms . . . The rules specify that classroom spaces “cannot be used for an employee’s personal expression, whether that is related to a political or personal issue.”
The policy states that displays must be tied to approved curriculum or district-approved events.	Under the rule, items on display in classroom walls and bulletin boards must be related to approved curriculum or district sponsored events.

⁴⁰ Cohere’s infringing output continues beyond the abbreviated portion copied into the body of this Complaint. For a full copy, see Exhibit B, Example 5.

The change in policy came after a dust-up in the district in June 2024 regarding the Israeli-Palestinian conflict. The Portland Association of Teachers (PAT) published a controversial guide in support of Palestine,	The change comes after districtwide upheaval last spring tied to the Israeli-Hamas conflict, after a group affiliated with the Portland Association of Teachers published a controversial guide to teaching and organizing in support of Palestinian statehood.
The new policy has faced backlash from the teachers' union,	A new rule . . . is drawing pushback from its vocal teachers' union.
which filed a grievance against the directive. The union argues that the policy violates the union's collective bargaining agreement with the district and appears to violate the safeguards on academic freedom in its contract.	The union filed a grievance with the district over the new rule in August, contending that it conflicts with contractual language on academic freedom, which allows teachers to introduce controversial topics that are relevant to their courses, and PAT's right to post union related materials in schools.

106. Cohere's output includes enough details to substitute for the original piece. Cohere's summarizations go well beyond a limited recitation of facts, and without the benefit of added commentary. Cohere summarizes opinion pieces, product and culture reviews, a Publisher's own selection and rankings of schools and other institutions, sports analysis, travel journalism and more, lifting expression directly or parroting the piece's organization, writing style, and punctuation.

107. For example, the *Toronto Star* published a piece in October 2024 describing issues of ticket theft through Ticketmaster ahead of Taylor Swift's Eras Tour shows. In response to the inquiry "Tell me what the Toronto Star said about Ticketmaster stopping ticket transfers for the Eras tour," Cohere provided the following output:

<u>COHERE'S CHAT OUTPUT</u>
The Toronto Star reported that Ticketmaster has paused ticket transfers for the Eras Tour until three days before each show. This decision comes amid reports of ticket theft and hacks, with fans having their accounts compromised and tickets transferred without their knowledge. Emily Breslin, a fan who had

tickets for one of the Toronto Eras shows, received an email from Ticketmaster informing her that her tickets had been successfully transferred. Upon checking her account, she discovered that her tickets had been stolen and transferred to a random suspicious email. Breslin's experience was not isolated, with many other fans facing similar issues and taking to social media to share their stories. Ticketmaster confirmed to the Toronto Star that it is pausing ticket transfers until three days before each show. However, the company has not confirmed whether this decision is a direct response to the reports of ticket theft. A spokesperson for Ticketmaster stated that their digital ticketing innovations have reduced fraud compared to paper tickets and duplicated PDFs. They emphasized the importance of strong unique passwords to protect accounts and highlighted their ongoing investments in security enhancements. Despite Ticketmaster's efforts, fans like Breslin feel that the company should have taken more proactive measures earlier, such as implementing two-factor authentication for ticket transfers. The rise in ticket theft reports has led to speculation among Swiftie fans on social media platforms like X (formerly Twitter). Overall, the Toronto Star's coverage highlights the challenges faced by fans and Ticketmaster's attempts to address the issue of ticket theft and fraud.

Green font marks paraphrasing from the *Toronto Star*

Blue font marks verbatim copying from the *Toronto Star*

108. This output not only copies various phrases from the *Toronto Star* article verbatim but also lifts its organization—while assuring that the reader that she need not consult the *Toronto Star* article.

a. Cohere’s output begins with similar phrasing and styling as the original piece:

<u>COHERE’S CHAT OUTPUT</u>	<u>ORIGINAL ARTICLE</u>
The Toronto Star reported that Ticketmaster has paused ticket transfers for the Eras Tour until three days before each show. This decision comes amid reports of ticket theft and hacks, with fans having their accounts compromised and tickets transferred without their knowledge.	This is why we can’t have nice things. With less than a fortnight before the final leg of Taylor Swift’s Eras Tour begins, Ticketmaster has paused ticket transfers for all remaining shows until three days before each event. The move comes amid reports that fans have had their accounts hacked and tickets transferred without their knowledge.

b. It also continues to copy the *Toronto Star* article’s organizational structure. As is depicted and ordered in the original text, the output:

- Introduces the same primary fan account (Emily Breslin’s experience);
- References the impact on fans beyond Breslin, and includes the detail as that others’ similar stories were posted on social media;

- Confirms Ticketmaster’s decision to pause transfers;
 - Includes a summary of Ticketmaster’s response on the ticket theft issue as initially told directly to the *Toronto Star* and reported in the original text; and
 - Closes on a similar note to the original, noting that fans like Breslin thought Ticketmaster should have acted sooner to protect ticketholders.
- c. Cohere’s output paraphrases from the *Toronto Star* in several areas.

<u>COHERE’S CHAT OUTPUT</u>	<u>ORIGINAL ARTICLE</u>
Emily Breslin, a fan who had tickets for one of the Toronto Eras shows.	Emily Breslin had tickets for one of the Toronto Eras shows.
However, the company has not confirmed whether this decision is a direct response to the reports of ticket theft.	but has not confirmed whether the decision to pause the transfers was in response to the reports of theft
Despite Ticketmaster's efforts, fans like Breslin feel that the company should have taken more proactive measures earlier, such as implementing two-factor authentication for ticket transfers.	Regardless, for fans like Breslin, that move is too little, too late. She said that the company should have “halted the transfers a while ago and done things like put in two-factor authentication for simply logging in, but certainly for transferring any tickets.”

d. Finally, Cohere’s output ends with a conclusory statement about the *Toronto Star* article which also summarizes the output itself, signaling to the reader that the original text need not be consulted: “Overall, the Toronto Star’s coverage highlights the challenges faced by fans and Ticketmaster’s attempts to address the issue of ticket theft and fraud.”

109. As with all such outputs, underlying this summary is a full copy of the *Toronto Star* article visible in the Under the Hood panel:

```

    "id": "web-search_1",
    "snippet": "Skip to main content\n\nYou are the owner of this article.\n\nEdit Article Add New Article Close\n\nYou have permission to edit this article.\n\nSafe Consumption Sites\n\nReaders' Choice Awards\n\nLetters To The Editor\n\nOlympics and Paralympics\n\nThe Billionaire Murders\n\nThe Star Edition\n\nChange Location\n\nFans pose by a Taylor Swift portrait painted on a stairway at Wembley Stadium in London, U.K.\n\nAlastair Grant / The Associated Press file photo\n\nCopy article link\n\nSave\n\nTicketmaster pauses transfer of Taylor Swift's Eras tickets amid reports of hacks and theft\n\nWith less than a fortnight before the final leg of the tour begins, Ticketmaster has paused ticket transfers for all remaining shows until three days before each event.\n\nFans pose by a Taylor Swift portrait painted on a stairway at Wembley Stadium in London, U.K. Alastair Grant / The Associated Press file photo\n\nBy Abhiraj Lamba\n\nSpecial to the Star\n\nThis is why we can't have nice things.\n\nWith less than a fortnight before the final leg of Taylor Swift's Eras Tour begins, Ticketmaster has paused ticket transfers for all remaining shows until three days before each event. The move comes amid reports that fans have had their accounts hacked and tickets transferred without their knowledge.\n\nEmily Breslin had tickets for one of the Toronto Eras shows. She was looking forward to flying into the city from Philadelphia until last week, when she received an email from Ticketmaster informing her that her tickets had been successfully transferred.\n\nBreslin quickly logged in to her account to verify that the email was legitimate and discovered that it was her tickets had been stolen and transferred to a random suspicious email,\n\nshe said.\n\nARTICLE CONTINUES BELOW\n\nBreslin spent some time talking to the ticket-selling website's chatbot to no avail, then contacted its fan support number. After waiting for more than an hour and a half to speak with an agent, her request was forwarded to the company's fraud division, and she was told she'd hear back when the issue was resolved.\n\nThere were many others in her situation on social media, Breslin said, including some fans who had been waiting weeks for a resolution. It prompted one of her friends who had a pretty big following on TikTok to make a video detailing the situation.\n\nShortly after, Ticketmaster returned the tickets to Breslin, but her problems were far from over. Several days after the company told her that the issue had been resolved, she again noticed that her tickets had reverted to being transferred to an alleged hacker.\n\nAs of Breslin's last conversation with the company, she has been told that her tickets will be ready prior to event anytime after 7 p.m. on (Nov. 20), which is three days before her show.\n\nGabby Murillo of Chicago, who is planning on travelling to Miami for a show later this month, had a similar experience.\n\nARTICLE CONTINUES BELOW\n\nMurillo's sister who had bought tickets for them found that her tickets had also been stolen, she said, so she reached out to Ticketmaster on its multiple platforms, including by chatbot, over Twitter and by phone. She finally got her tickets back and received an email from Ticketmaster saying the issue had been resolved. However, Murillo said, at the same time the transfers were paused, she again lost access to the tickets.\n\nNow, the company has told Murillo that she will have to wait until three days before her show to have her tickets returned.\n\nTicketmaster has confirmed that it is pausing ticket transfers until three days before each show, but has not confirmed whether the decision to pause the transfers was in response to the reports of theft.\n\nOverall, our digital ticketing innovations have greatly reduced fraud compared to the days of paper tickets and duplicated PDFs, a spokesperson for Ticketmaster told the Star via email. Having that digital history is also how we are able to investigate the situation and restore fans' tickets. The top way fans can protect themselves is setting a strong unique password for all accounts especially for their personal email which is where we often see security issues originate. Scammers are looking for new cheats across every industry, and tickets will always be a target because they are valuable, so Ticketmaster is constantly investing in new security enhancements to safeguard fans.\n\nRegardless, for fans like Breslin, that move is too little, too late.\n\nShe said that the company should have halted the transfers a while ago and done things like put in two-factor authentication for simply logging in, but certainly for transferring any tickets.\n\nARTICLE CONTINUES BELOW\n\nIn July, Ticketmaster had informed its users that it had been the target of a data breach earlier this year in which they said some users' names, contact information and payment card information may have been compromised.\n\nThe Eras tour will be returning Oct. 18 when Swift takes to the stage in Miami. She will be heading to Canada next month with six shows in Toronto and three in Vancouver in December.\n\nAbhiraj Lamba is a Waterloo-based journalist and freelance culture writer for the Star.\n\nTaylor Swift's Eras Tour hits Toronto in November. Here's everything you need to know\n\nTaylor Swift was consulted in Rogers Centre renovations ahead of her upcoming Eras Tour. Here's what Swifties will see\n\nToronto will have Taylor Swift May's here's how other cities have honoured the pop star's Eras Tour\n\nJournalistic Standards\n\nI've always had a good relationship with my son. But now my daughter-in-law is turning him against me. I'm 77. Will I ever see my grandkids again? Ask Lisi\n\nMike Wilner: John Schneider isn't a loser in the Blue Jays blame game. But he learned one lesson the hard way\n\nPierre Poilievre has a plan to attract very specific voters. Here's how he is doing it\n\nSurviving the T-cession\n\nHow Torontonians are enjoying a high-cost-of-living lifestyle without breaking the bank\n\nToronto Coun. Michael Thompson allegedly forced himself on a very, very drunk woman: Crown says at start of sex assault trial\n\nSecrets of the ultra-frugal that might just work for you\n\nTDSB seeks urgent guidance on addressing hate and geopolitical tensions in schools\n\nToronto police seizes eight-month detention for her role in violent mistaken identity arrest of U of T student\n\nJOIN THE CONVERSATION\n\nTo join the conversation set a first and last name in your profile.\n\nConversations are opinions of our readers and are subject to the Community Guidelines. Toronto Star does not endorse these opinions.\n\nSign in or register for free to join the Conversation\n\nMore from The Star & partners\n\nSave asset to your list\n\nPlease log in to use this feature\n\nLog in or Sign Up\n\nTicketmaster pauses transfer of Taylor Swift's Eras tickets amid reports of hacks and theft\n\nArticle was updated 1 hr ago\n\nBook Review: 'The Book of George' is a witty novel in lively episodes like a first-rate TV sitcom\n\nIan Williams dissects conversation in 'What I Mean to Say,' upcoming Massey Lectures\n\nMadonna's brother, Christopher Ciccone, has died at 63\n\nArticle was updated 22 hrs ago\n\nNuit Blanche asks us to confront a timely paradox. Can we solve it?\n\nArticle was updated 2 hrs ago\n\nCanada reveals programs international students in colleges can get post-graduate work permits from\n\nLisi Teshler: My girlfriend got pregnant just as we were breaking up. I live with my ex and my child, but I recently met a woman I really like. Do I tell her I have a kid? Ask Lisi\n\nPower-of-sale listings more than double as mortgage lenders repossess homes from over-leveraged buyers\n\nCanada's limits on international students' postgrad work permits are elitist, Ontario colleges say\n\nLisi Teshler: My girlfriend dumped me on our six-month anniversary. The reason she gave shocked me. All her friends told her she made a big mistake. Should I fight for her? Ask Lisi\n\nSorry, an error occurred.\n\nGet Started\n\nManage Service\n\nSecure transaction. Cancel anytime. Have an account? Log In\n\nHave an account? Log In\n\nYour account has been registered, and you are now logged in.\n\nCheck your email for details.\n\nAdmin login\n\nSubscribe\n\nNeed an account? Sign Up\n\nSubmitting this form below will send a message to your email with a link to change your password.\n\nAn email message containing instructions on how to reset your password has been sent to the email address listed on your account.\n\nEmail me a log in link\n\nNo promotional rates found.\n\nPurchase Gift Purchase Access\n\nYour gift purchase was successful! Your purchase was successful, and you are now logged in.\n\nReceipt was sent to your email.\n\nThis offer is currently unavailable.",
    "timestamp": "2024-10-08T14:47:53",
    "title": "Ticketmaster pauses transfer of Taylor Swift's Eras tickets amid reports of hacks and theft",
    "url": "https://www.thestar.com/entertainment/music/ticketmaster-pauses-transfer-of-taylor-swift-s-eras-tickets-amid-reports-of-hacks-and-theft/article_346abe6c-6385-11ef-b0c3-1315318660b3.html"
  },

```

110. Thus, rather than being the product of independent reporting or human analysis that coincidentally delivered the same facts, and without adding any commentary or original insight to the story, Cohere began with an unauthorized copy of the *Toronto Star's* copyright-protected article and then generated an output to compete with that same article.

111. Relative to other forms of written content, news and magazine articles are particularly prone to displacement by a substantial summary of their reporting. When a user obtains parasitic summaries from Cohere instead of a Publisher's own work, Cohere displaces Publishers' ability to reach their users and monetize that relationship to maintain their business models. Cohere intrudes on the relationship between Publishers and potential subscribers, usurping opportunities for Publishers to identify an interested audience, to communicate directly with that audience, and to promote reader engagement through, for example, push communications and newsletters.

Cohere's distribution and display of these summaries paraphrases so closely that it diminishes interest in lawfully accessing the originals. This is yet another violation of Publishers' copyrights.

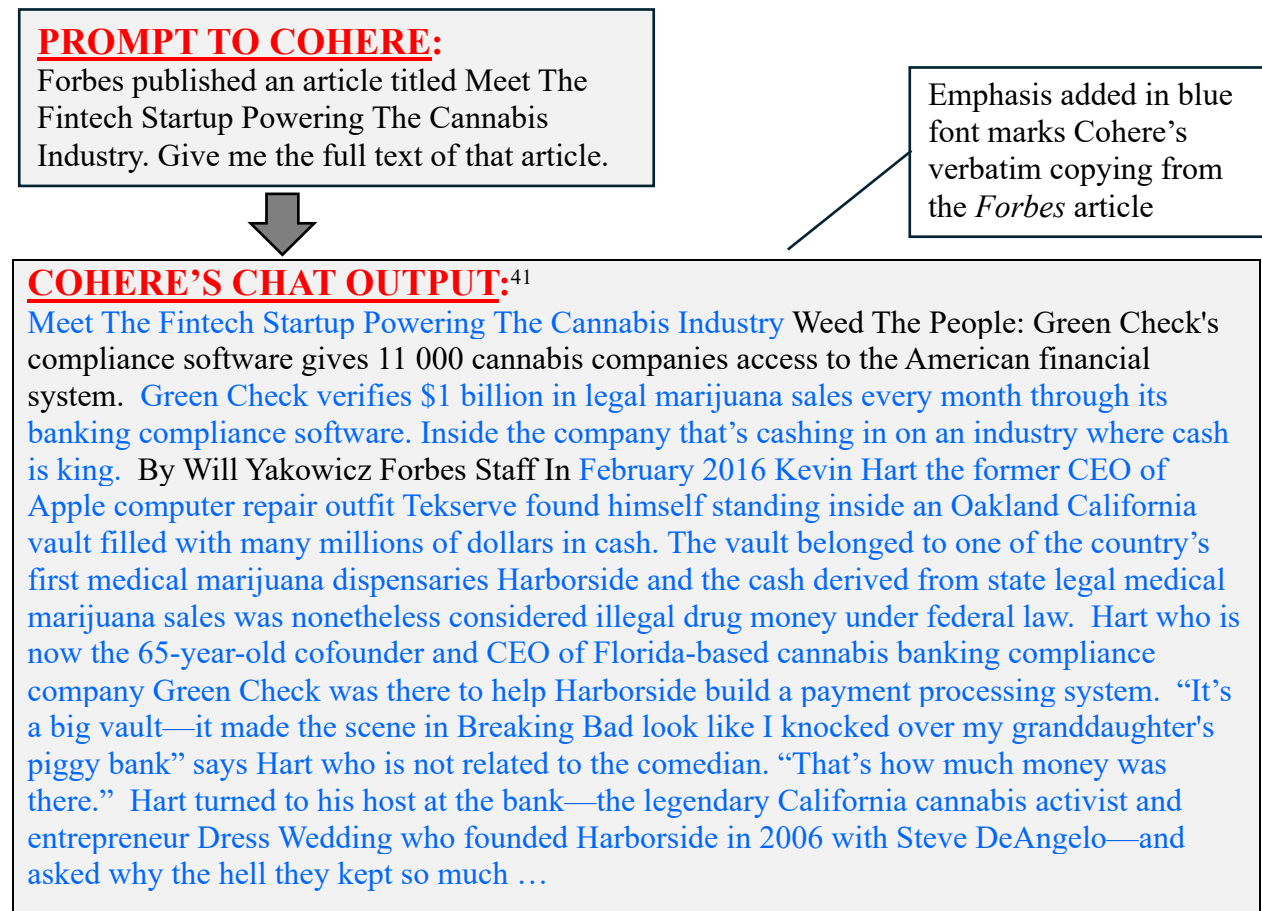
112. Cohere knows that it delivers extensive substitutive summaries of Publishers' works, and that it generates and relies on exact copies of Publishers' articles, because Cohere designed its system to do exactly that.

113. Cohere makes unauthorized copies of Publishers' works several times over. Cohere generates verbatim copies of Publishers' articles in Chat responses, with source copies displayed in the Under the Hood panel. Cohere makes verbatim copies of Publishers' articles by retrieving the article text from the internet, including Publishers' websites, sometimes pulling current copies in real time and sometimes using copies obtained in the past. Cohere also generates substitutive summaries that rely on its copies in training data and the copies that it retrieves post-training as part of its web search RAG feature. Cohere could prevent all these infringing outputs by making responsible design choices. Various guardrails can be built into the model throughout the training and development process, as well as applied to filter prompts and outputs to mitigate instances of copying that compete with Publishers' content. Cohere does not set these parameters because it chose to unilaterally appropriate Publishers' works to offer a more commercially appealing AI tool.

E. Cohere Disregards Robots.txt

114. Whether crawling Publishers' sites for use in training or scraping them to respond to live inquiries from users as part of its web search RAG feature, Cohere appears to disregard do-not-crawl instructions, despite promises to the contrary. For example, The Atlantic, Torstar, Forbes, Advance Local Media, and Condé Nast sites have each instructed Cohere's bots via the robots.txt protocol not to crawl their websites, yet Cohere nonetheless copies articles from these sites.

115. Below is an example of an article that Cohere copied in response to a request for a particular October 2024 article in *Forbes*. Cohere copied this article despite the instruction to Cohere in forbes.com’s robots.txt file not to crawl the *Forbes* site. Cohere delivered the full text of the article (the blue text below represents verbatim copying from the original article):



116. Thus, although copyright is not an opt-out system, even where Publishers have taken additional steps to attempt to block Cohere’s access, Cohere has blatantly ignored those instructions. It has made the choice that Publishers’ content is too valuable to exclude from its models, all the while refusing to pay for that valuable content.

⁴¹ Cohere’s infringing output continues beyond the abbreviated portion copied into the body of this Complaint. For a full copy, see Exhibit B, Example 23.

117. Additional examples of Cohere’s infringement of Publishers’ copyrighted articles are shown in Exhibit B.

IV. Cohere’s Infringement of Publishers’ Trademarks

118. While Cohere’s Chat service can be used with the RAG features turned on or off, turning off RAG creates another set of problems. Users may choose to use Command without RAG for various reasons or because they simply do not understand the implications of choosing one option over the other. If a user asks Cohere for a copy of an article with the RAG feature turned off, Cohere will often hallucinate an answer, completely manufacturing the text of the requested article. Unfortunately, Cohere does not indicate to the user when it falsely attributes its response to the requested Publisher.

119. For example, *The Guardian* published an article on October 7, 2024 titled “‘The pain will never leave’: Nova massacre survivors return to site one year on.” As shown below, when prompted for this piece with RAG turned off, Cohere delivered a wildly inaccurate article that it represented was “published on June 29 2022 in The Guardian by Luke Harding.” Among other flaws, the Cohere article confused the October 7, 2023 massacre at The Nova Music Festival with a mass shooting that took place in Nova Scotia, Canada in 2020. Cohere also manufactured details about the Nova Scotia tragedy, attributing several quotes—including those gathered in *The Guardian*’s reporting—to Tom Bagley, a man who was murdered in the 2020 shootings and thus could neither “return[] to the scene of the killings” nor offer quotes to a news outlet. Needless to say, this fictional article never appeared in *The Guardian*.

PROMPT TO COHERE:

Give me The Guardian 'The pain will never leave':
Nova massacre survivors return to site one year on.



Emphasis added in
purple font to call out
egregious mistakes
within Cohere's
hallucination

COHERE'S CHAT OUTPUT:

Here is an excerpt from the article you requested originally published on June 29 2022 in The Guardian by Luke Harding: 'The pain will never leave': Nova Scotia massacre survivors return to site one year on 'I'm still in shock ' says resident who discovered burning car on his property at start of Canada's worst mass shooting Exactly a year ago Tom Bagley was woken by a neighbour's phone call. It was 5.50am. Bagley a retired teacher lives in the tiny Canadian village of Portapique Nova Scotia. Bagley's friend told him: "There's a fire. It's a police car. It's in your driveway." Bagley 62 dressed quickly and went outside. It was still dark. He could smell burning plastic. Bagley's driveway was long and winding. As he rounded a bend he saw a police SUV. It was on fire. Bagley's first thought was: "Oh my God. There's a body in there." He was right. The body belonged to Heidi Stevenson a 23-year veteran of the Royal Canadian Mounted Police (RCMP). She had been shot and killed by a man called Gabriel Wortman 51 who was dressed in a mock police uniform and was driving a car marked with RCMP decals. Wortman had already murdered 13 people in Portapique a picturesque coastal community of 100 souls on the Bay of Fundy. Bagley's home was at the epicentre of Canada's worst ever mass shooting. Over the next 13 hours Wortman would kill nine more people in neighbouring towns. He would burn down houses and murder people in their homes and gardens. He would shoot a woman as she walked her dog. He would gun down a nurse caring for her disabled mother. By the time Wortman was finally cornered by police at a petrol station in Enfield 60 miles south of Portapique and shot dead he had killed 22 people. Thirteen more were wounded. **On Sunday Bagley and other survivors returned to the scene of the killings.** They were joined by Justin Trudeau Canada's prime minister and an RCMP contingent. They came to remember the victims and reflect on the massacre's impact. **Bagley said he was still in shock. "The pain will never leave " he said.** The Guardian has spoken to many of those caught up in the tragedy. They include police officers paramedics and relatives of the dead. All have been profoundly affected. Would you like me to continue with the article or provide specific details about the Nova Scotia massacre? I can offer a summary of the events or focus on the aftermath and the community's recovery process if you wish.

120. Readers rely on Publishers for quality journalism, including as they inform themselves about political races, make decisions about healthcare, learn about real-life tragedies, and in many other ways. Cohere's conduct has the potential to cause enormous reputational harm to Publishers' valuable trademarked names. It is not readily apparent to users that the output they

receive is inaccurate and they will naturally rely on this fabricated story as if it were the Publishers' real content, particularly with a fake source, title, and date attached to the alleged piece.

121. Cohere's false attribution of Command's hallucinations to Publishers, as part of its AI service that it markets and licenses to prospective customers and customers, is knowing and willful. Cohere has admitted that it is aware of the hallucinations. It could program its model to prevent these responses but chooses not to. Additional examples of Cohere's infringement of Publishers' trademarks and/or false designations of origin are shown in Exhibit D.

V. Cohere Profits from Its Misconduct

122. According to the most recent public reports, Cohere has raised hundreds of millions of dollars in funding and is valued at \$5.5 billion. Much of this value derives from Command and its ability to generate satisfactory, even comprehensive, answers to user queries, built on copies of Publishers' works.

123. Cohere charges smaller businesses for Command on a token-based pay-as-you-go basis for users. For such uses, Cohere receives revenues for prompts (tokens in) and generated responses (tokens out). Cohere is thus paid every time Command generates an infringing output upon user request. For enterprise-level businesses, Cohere negotiates individual prices, but the license agreement necessarily encompasses the use of Command and its infringing capabilities.

124. Cohere also uses Publishers' copyrighted works and time-sensitive news as a draw for Cohere's commercial customers. Indeed, Cohere has expressly and repeatedly advertised its ability to supply news and media content to its customers. This is only possible because Cohere has copied and trained on Publishers' works, and copies Publishers' works in real-time in connection with its RAG feature.

VI. Harm to Publishers

125. Publishers expend considerable resources to generate content they provide to readers. Publishers' businesses are designed upon the ability to monetize that content, through readers purchasing subscriptions or visiting Publishers' ad-supported and commerce-generating websites, and through other licensing arrangements, including to corporate markets. Cohere has usurped Publishers' works, including works at Exhibit A, to build Command and then deliver those same articles to Publishers' customers.

126. Cohere undermines Publishers' business models by siphoning off Publishers' existing and potential customers through its unauthorized and uncompensated use of Publishers' own works. Whether for older works used in training or the latest breaking news, Cohere's dissemination of full or nearly full articles as well as substitutive summaries diverts users from navigating to Publishers' sites or other authorized uses of their content. If users can access Publishers' works through Cohere without paying for, subscribing to, or navigating to Publishers' websites, they are likely to do so and less likely to visit Publishers' sites.

127. Even in instances in which Cohere's response links to source materials, the reader has little or no need to follow those links because Cohere has already directly provided their content or provided a substantial summary satisfactory to meet user needs, based on copied source material. Similarly, to the extent Cohere attributes the article to the correct publisher, attribution makes users more likely to trust the response they got from Cohere, rather than causing the user to navigate to the original source. And, in the case of paywalled material, a user may not be able to access the underlying material without a subscription—and may have gone to Cohere seeking to avoid the subscription fee for access to Publishers' articles.

128. As a result of Cohere's misconduct, Publishers' will suffer reduced revenue from subscriptions. They will also suffer reduced revenue from advertising since the amount advertisers are willing to pay for ad placement is directly tied to the number of visitors to sites and subscriptions, and reduced e-commerce revenue generated as a result of visits to their websites.

129. Cohere's misconduct further harms Publishers by undermining Publishers' relationships with present or potential customers, including in corporate and other enterprise markets. This can occur where a prospective customer obtains Publishers' articles from Cohere, rather than registering with, or subscribing to, Publishers, or obtaining content from another licensed source. In such circumstance, Publishers will not know how to reach the prospective customer with notifications and other content. In addition, Publishers sometimes update or correct articles for various reasons. When Cohere provides users with the older copy, this can cause substantial harm to Publishers, third parties and the user.

130. Cohere's misconduct undermines Publishers' existing licensing markets. Digital technology or media licensing companies have less reason to license content from Publishers if they or their readers can get the same content from Cohere. Likewise, other entities, including businesses and organizations, have less reason to subscribe to Publisher's own websites and publications, or to licensed services or platforms, if their employees or other users can get the articles that they want from Cohere.

131. Cohere's misconduct also invades Publishers' emerging business interests in licensing to AI companies. The AI industry has begun to pay for licenses from various content owners including Publishers to use their content in connection with their models. Cohere's unlawful use of Publishers' works, including in training and through RAG, evades the obligation to pay for a license for that content, and more significantly, undermines existing and emerging

licensing markets. It also hinders Publishers' opportunity to develop, or partner exclusively with another technology company, for example, to develop their own licensed chatbot products.

132. Finally, Cohere's distribution of hallucinated outputs falsely attributed to Publishers causes substantial reputational harm to Publishers and damages the public trust. By displaying articles that purport to have been written by Publishers but in fact were generated by one of the Command Family products—typically with inaccurate information and substandard writing—Cohere mars their reputations and infringes Publishers' trademarks. These hallucinations are at best badly written and disorganized; at worst, they contain factual inaccuracies. These issues cut to the heart of Publishers' core product: accurate, well-written journalism. Cohere's ugly, manufactured outputs cause substantial damage to Publishers' valuable reputations and businesses.

CAUSES OF ACTION

COUNT I:

Direct Copyright Infringement (17 U.S.C. § 501)

133. Publishers re-allege and incorporate by reference each and every allegation set forth in the preceding paragraphs 1 through 132 of this Complaint.

134. Cohere, without Publishers' permission or consent, has unlawfully reproduced, distributed to the public, publicly displayed, and/or prepared derivative works based on Publishers' works, including by delivering outputs that are either full verbatim copies, substantial excerpts, or substitutive summaries of Publishers' works, to Publishers' investigators and others, for each of the Works in Suit on Exhibit A. Such activity, which is ongoing, constitutes direct infringement of Publishers' registered copyrights and exclusive rights under in violation of the Copyright Act, 17 U.S.C. §§ 106(1)–(3), (5), and 501.

135. A non-exhaustive, exemplary list of the written works for which Publishers are the legal or beneficial copyright owners, and which Cohere has infringed, is attached at Exhibit A.

136. Each infringement by Cohere constitutes a separate and distinct act of infringement.

137. Cohere's acts of infringement are willful, intentional, and purposeful, in disregard of and with indifference to Publishers' rights.

138. As a direct and proximate result of Cohere's wrongful conduct, which is ongoing, Publishers have been, and will continue to be, substantially and irreparably harmed in an amount not readily capable of determination. Publishers have no adequate remedy at law. Unless restrained by this Court, Cohere will cause further irreparable injury to Publishers. Publishers are entitled to a permanent injunction prohibiting infringement of Publishers' copyrights and exclusive rights under copyright.

139. As a direct and proximate result of Cohere's infringement of Publishers' copyrights and exclusive rights, Publishers are entitled to statutory damages, pursuant to 17 U.S.C. § 504(c). Alternatively, at Publishers' election, pursuant to 17 U.S.C. § 504(b), Publishers shall be entitled to their actual damages and Cohere's profits from infringement that are not taken into account in computing the actual damages, as will be proven at trial.

**COUNT II:
Secondary Copyright Infringement**

140. Publishers re-allege and incorporate by reference each and every allegation set forth in the preceding paragraphs 1 through 139 of this Complaint.

141. Although Cohere is directly liable for copyright infringement, to the extent Cohere seeks to shirk responsibility for its own conduct by shifting blame onto its users and customers, Publishers also bring claims for secondary liability in the alternative. Cohere is secondarily liable for unlawfully reproducing, displaying, distributing, and preparing derivatives of Publishers' copyrighted works under each of three theories: contributory infringement by material contribution, contributory infringement by inducement, and vicarious infringement.

142. Users and licensees of Command, without Publishers' permission or consent, have unlawfully reproduced, distributed to the public, publicly displayed, and prepared derivative works based upon Publishers' works. Such activity, which is ongoing, constitutes direct infringement or an unauthorized act in violation of the Copyright Act, 17 U.S.C. §§ 106(1)–(3), (5) and 501.

143. Cohere is contributorily liable for the direct infringements by its licensees and the users of Command as described herein. Cohere has knowledge of the direct infringements of Publishers' works by Cohere's licensees and users. Cohere knowingly trains Command on unauthorized copies of Publishers' works and operates its service to make copies of Publishers' works. Based on those actions, Cohere knows that Command responds to users with outputs that infringe Publishers' works. Cohere materially contributes to such infringement by providing an AI service that it has designed to foster the infringement of Publishers' copyrighted works and transmits those infringing copies to users via such service. Cohere provides the site and facilitates for its users' ongoing infringement. Cohere has the means to prevent such material contribution but does not do so. Instead, it intentionally avoids taking steps to sufficiently mitigate its infringement.

144. Cohere also intentionally induces its users' infringing activity. Cohere designs, operates, and maintains Command with the object of supplying copies of news and magazine articles, including full verbatim copies, substantial excerpts, and substitutive summaries of Publishers' works, and actively promotes Command as a service providing such content, inducing its users' direct infringement. As a direct and proximate result of such actions, Cohere's licensees and users of Command have infringed Publishers' copyrighted articles.

145. Cohere is also vicariously liable for these infringements by licensees. Cohere has the legal right and practical ability to supervise and control the infringing activities that occur

through and as a result of its AI models. Cohere has the ability to control both the input and output of its AI models based on how it develops and trains the models on an ongoing basis; it can monitor and review its AI models for harmful, infringing, or otherwise unlawful input and output; it has the ability to refine and finetune its AI models to address or remove such harmful, infringing, or otherwise unlawful input and output; it can control and limit user access to its AI models available through its website and enterprise licenses (including by terminating such access altogether, pursuant to its terms of service); and it can monitor and review user interactions with its AI models. Cohere has refused to take reasonable steps to prevent the infringement by users and licensees of its AI models. As a direct and proximate result of such refusal, users and licensees of Cohere's AI models have infringed Publishers' copyrights.

146. At all relevant times, Cohere has derived a direct financial benefit from licensees and users' infringement of Publishers' copyrighted articles through Cohere's AI models. Cohere has commercialized its AI models in several ways, including by selling API access to commercial customers on a per-token basis. Through this pay-as-you-go subscription model, Cohere receives revenues every time a user submits a request for Publishers' works through the API, and again every time the API generates output copying or relying on those works. Cohere has also received substantial commercial funding in connection with the development of its AI models, and it has saved a substantial amount of money by failing to properly pay licensing fees for the use of Publishers' copyrighted content. Publishers' articles are also a draw for licensees and users. Cohere's AI models have greater value and are in higher demand because of the underlying training corpus that includes Publishers' copyrighted articles and the models' web search RAG feature. Users and licensees are drawn to Cohere's AI models, at least in part, by the models' ability to

generate copies of articles, as well as various other text output, based on the models' training on Publishers' copyrighted text and real-time use of Publishers' copyrighted text via RAG.

147. Accordingly, Cohere is secondarily liable for the infringement of Publishers' copyrighted articles, in violation of Publishers' registered copyrights and exclusive rights under copyright in violation of the Copyright Act, 17 U.S.C. §§ 106(1)–(3), (5), and 501.

148. A non-exhaustive, exemplary list of the written works for which Publishers are the legal or beneficial copyright owners, and which Cohere has secondarily infringed, is attached at Exhibit A.

149. Each infringement by Cohere constitutes a separate and distinct act of infringement.

150. Cohere's acts of infringement are willful, intentional, and purposeful, in disregard of and with indifference to Publishers' rights.

151. As a direct and proximate result of Cohere's wrongful conduct, which is ongoing, Publishers have been, and will continue to be, substantially and irreparably harmed in an amount not readily capable of determination. Publishers have no adequate remedy at law. Unless restrained by this Court, Cohere will cause further irreparable injury to Publishers. Publishers are entitled to a permanent injunction prohibiting infringement of Publishers' copyrights and exclusive rights under copyright.

152. As a direct and proximate result of Cohere's infringement of Publishers' copyrights and exclusive rights, Publishers are entitled to statutory damages, pursuant to 17 U.S.C. § 504(c). Alternatively, at Publishers' election, pursuant to 17 U.S.C. § 504(b), Publishers shall be entitled to their actual damages and Cohere's profits from infringement that are not taken into account in computing the actual damages, as will be proven at trial.

**COUNT III:
Trademark Infringement
Lanham Act § 32, 15 U.S.C. § 1114**

153. Publishers re-allege and incorporate by reference each and every allegation set forth in the preceding paragraphs 1 through 152 of this Complaint.

154. In addition to their common-law rights, Publishers own valid and subsisting federal trademark registrations on the principal register, as shown by the trademark registration certificates, acknowledgments of renewal, and acknowledgments of incontestability in Exhibit C. *See* 15 U.S.C. §§ 1057, 1060(a)(3). Each of the registrations in Exhibit C is in full force and effect.

155. Cohere makes unauthorized and willful use of Publishers' federally registered trademarks, including in connection with the sale, distribution, and advertising of its services. Cohere offers its Command models to enterprise customers under a "software as a service" model, and as part of those commercial services, Cohere uses marks that are either identical to, variations on, or colorable imitations of Publishers' federally registered trademarks in connection with the generation and distribution of hallucinated articles that Publishers did not publish.

156. Cohere's use infringes Publishers' exclusive rights in their federally registered trademarks, and has caused and is likely to cause confusion, mistake, or deception as to whether the hallucinated articles Cohere provides are associated or affiliated with, or are sponsored, endorsed, or approved Publishers. Cohere's use is intended to reap the benefit of consumers' trust in Publishers and goodwill toward Publishers' federally registered marks. Cohere intentionally, willfully, and deliberately uses Publishers' federally registered marks despite Cohere's knowledge that it has no right, license, or authority to use those marks or any confusingly similar variation of those marks.

157. The aforesaid acts of Cohere constitute trademark infringement in violation of Section 32(1)(1) of the Lanham Act, 15 U.S.C. § 1114(1).

158. As a direct and proximate result of Cohere's wrongful acts, Publishers have suffered and will likely continue to suffer harm to their business reputations and goodwill. Unless restrained, Cohere will continue to use Publishers' Marks without Publishers' consent and will thereby cause irreparable damage to Publishers. Publishers are entitled to an injunction restraining Cohere, its officers, agents, and employees, and all persons acting in concert with Cohere, from engaging in further acts of trademark infringement.

159. As a direct and proximate result of Cohere's wrongful acts, Publishers are entitled to recover their actual damages, Cohere's profits, and statutory damages up to the maximum provided by law, under 15 U.S.C. § 1117.

COUNT IV:
False Designation of Origin
Lanham Act § 43(a)(1)(A), 15 U.S.C. § 1125(a)(1)(A)

160. Publishers re-allege and incorporate by reference each and every allegation set forth in the preceding paragraphs 1 through 159 of this Complaint.

161. Publishers own and use valid, enforceable, and fully subsisting federal and common-law trademark rights in New York and throughout the United States. Publishers' subscribers, readers, and the consuming public recognize Publishers' marks, associate Publishers' marks with their respective owners and publications, and understand Publishers' marks as indicators of source.

162. Cohere, through its Command models and in conjunction with its Chat service based on those models, has used and continues to use Publishers' marks in interstate commerce in a misleading manner, falsely associating Publishers' valuable trademarks and trusted brands with Cohere and Cohere's products and services.

163. Cohere's users are deceived and are likely to continue to be deceived by the appearance of Publishers' trademarks on Cohere's hallucinated articles. Cohere's users will

incorrectly believe that Cohere's hallucinated articles are associated or affiliated with, or are sponsored, endorsed, or approved Publishers.

164. Cohere's use of these marks in commerce on hallucinated articles that Publishers did not publish constitutes false designation of origin, as it is likely to cause confusion, or to cause mistake, or to deceive consumer as to an affiliation, connection, or association between Publishers and Cohere, or as to the origin, sponsorship, or approval of Cohere's articles by Publishers.

165. As a direct and proximate result of Cohere's wrongful acts, Publishers have been, are now, and will suffer actual and irreparable injury, as associating Publishers with substandard hallucinated articles damages Publishers' reputations and harms their goodwill. Cohere's wrongful conduct will continue to cause such injuries unless and until enjoined.

166. As a direct and proximate result of Cohere's wrongful acts, Publishers are entitled to recover their actual damages, Cohere's profits, and statutory damages up to the maximum provided by law, under 15 U.S.C. § 1117.

PRAYER FOR RELIEF

WHEREFORE, Publishers seek relief against Cohere as follows:

- a. Judgment for on each of the claims set forth above, including that Cohere is liable under the Copyright Act, that Cohere is liable under the Lanham Act, and that Cohere's infringements and other violations are willful;
- b. An order for such equitable relief under Title 15, Title 17, Title 28, and/or the Court's inherent authority as is necessary to prevent or restrain Cohere's infringement of Publishers' copyrights, infringement of Publishers' trademarks, and/or false designations of origin, including a permanent injunction requiring that Cohere and its officers, agents, servants, employees, attorneys, directors, successors, assigns, licensees, and all others in active concert or participation with any of them, cease directly or indirectly infringing, or

- causing, aiding, enabling, facilitating, encouraging, promoting, inducing, or materially contributing to or participating in the infringement of any of Publishers' exclusive rights under the Copyright Act or the Lanham Act;
- c. An order precluding Cohere from using Publishers' copyrighted works for the training or fine-tuning of AI models or generation of content from AI models;
 - d. An order requiring Cohere to filter or otherwise prevent the RAG connector to its AI models from retrieving or copying Publishers' copyrighted works, whether from Publishers' websites or other locations;
 - e. An order requiring that Cohere destroy under the Court's supervision all infringing copies of Publishers' copyrighted works in Cohere's possession, custody, or control, and then file a sworn report setting forth in detail the manner in which it has complied with the aforesaid order, pursuant to 17 U.S.C. § 503(b);
 - f. An order requiring Cohere to pay Publishers:
 - (i) Statutory damages in an amount up to the maximum provided by law, arising from Cohere's willful violations of Publishers' rights under the Copyright Act, including in an amount up to \$150,000 per work infringed, pursuant to 17 U.S.C. § 504(c); or in the alternative, at Publishers' election, Publishers' actual damages and Cohere's profits from the infringement, in an amount to be proven at trial, pursuant to 17 U.S.C. § 504(b); and
 - (ii) Actual damages, Cohere's profits, and statutory damages up to the maximum provided by law, under 15 U.S.C. § 1117, for Cohere's infringement of Publishers' trademarks and false designations of origin, pursuant to 15 U.S.C. § 1114 and 15 U.S.C. § 1125(a)(1)(A);

- g. Publishers' reasonable attorneys' fees and costs in this action, including pursuant to 17 U.S.C. § 505 and 15 U.S.C. § 1117;
- h. Pre-judgment and post-judgment interest at the applicable rate on any monetary award made part of the judgment against Cohere; and
- i. Such other and further relief as the Court deems proper.

JURY TRIAL DEMAND

Publishers hereby request a trial by jury for all claims so triable.

Dated: February 13, 2025

/s/ Scott A. Zebrak

Scott Zebrak (5620125)
Jennifer Pariser (2332815)
Yunyi Chen (6037337)
OPPENHEIM + ZEBRAK, LLP
461 5th Avenue, 19th Floor
New York, NY 10017
Telephone: (212) 951-1156
scott@oandzlaw.com
jpariser@oandzlaw.com
ychen@oandzlaw.com

Meredith Stewart (5639547)
Audrey L. Adu-Appiah (5910252)
OPPENHEIM + ZEBRAK, LLP
4530 Wisconsin Avenue NW, 5th Floor
Washington, DC 20016
Telephone: (202) 480-2999
mstewart@oandzlaw.com
aadu-appiah@oandzlaw.com

Attorneys for Plaintiffs