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

GRACENOTE MEDIA SERVICES, LLC, Plaintiff, v. OPENAI FOUNDATION (f/k/a OPENAI, INC.), OPENAI GROUP PBC, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO, LLC, OPENAI GLOBAL, LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC, Defendants.	Civil Action No. 1:26-cv-1947 **JURY TRIAL DEMANDED**

COMPLAINT

Plaintiff Gracenote Media Services, LLC ("Gracenote"), by and through its attorneys Susman Godfrey LLP, brings the following Complaint against Defendants OpenAI Foundation (f/k/a OpenAI, Inc.), OpenAI Group PBC, OpenAI GP, LLC, OpenAI, LLC, OpenAI OpCo, LLC, OpenAI Global, LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC (collectively "OpenAI" or "Defendants"), and alleges as follows:

I. NATURE OF THE ACTION

1. For over half a century, Gracenote has dedicated itself to helping media consumers—and the companies that serve them—navigate and discover media content. Relying on the hard work, professional judgment, editorial insight, and media expertise of hundreds of Gracenote editors, Gracenote has developed a powerful and original set of metadata products, including Gracenote's copyrighted Gracenote Programs Database and its data (collectively, "Gracenote Data"). Gracenote Data includes millions upon millions of narrative descriptions, original video descriptors, unique identifiers, and other program elements that have been carefully

written, edited, selected, and arranged by Gracenote editors. The scale, integrity, relational logic, and originality of Gracenote Data is tremendously valuable. It is valuable for Gracenote's core customers: the country's largest providers of media content distributors (cable, satellite, and streaming) who license Gracenote Data so that their own customers (content consumers) can better navigate the ever-expanding universe of media. And Gracenote Data is highly valuable for a new era of market-leading artificial intelligence ("AI") and machine learning ("ML") providers, who directly license Gracenote Data to train models and significantly improve the quality and accuracy of AI-generated content in consumer-facing AI products.

2. Without permission or compensation to Gracenote, Defendants have copied and used Gracenote Data to create and improve highly lucrative AI products like ChatGPT, powered by large-language models ("LLMs" or "models"). Defendants' AI products contain and generate exact copies of Gracenote Data, including outputting Gracenote Data's descriptive content verbatim. Defendants' AI products compete with and threaten Gracenote's core business with media content distributors and growing business involving providers of AI products themselves. Gracenote's media content distributor customers may use Defendants' AI products as (or to build) their own substitutive, competing media metadata products and platforms, all without permission or compensation to Gracenote.

3. Likewise, Defendants' copying and distribution of Gracenote Data through Defendants' AI products is harming Gracenote in the emerging market for data to train and/or improve the quality and accuracy of AI products—specifically, Defendants are eroding Gracenote's ability to license Gracenote Data to competing providers of AI products who need Gracenote Data to utilize content metadata effectively. Further, Defendants' unconstrained use of Gracenote Data in AI products runs roughshod over carefully negotiated use restrictions in such licenses. These licenses ensure that any AI products that utilize Gracenote Data do not siphon away

Gracenote's core base of media content distributor customers, as well as ensure that Gracenote Data is used properly in the marketplace to drive value and audiences back to the associated content.

4. Defendants could have paid Gracenote to license its valuable Gracenote Data.[1] Or they could have sought to train and ground their models only on information in the public domain. They did neither. Defendants instead improperly copied and used Gracenote Data to create their own commercially valuable AI products, all without paying a dime.

5. Defendants' actions violate Gracenote's exclusive rights to its creative works and intellectual labor under the Copyright Act and other applicable laws. Those laws protect Gracenote and millions of its original records from Defendants trampling on Gracenote's intellectual property with impunity.

6. To build OpenAI's LLMs, and develop, operate, and improve their associated commercial products like ChatGPT, Defendants have knowingly:

a. Copied Gracenote Data without authorization;

b. Used Gracenote Data to develop OpenAI's LLMs and develop, operate, and improve their LLM-based products and services, which OpenAI provides to third parties;

c. Used OpenAI's LLMs and LLM-based products and services to deliver infringing content to users of its products and online services;

d. Undermined and eroded Gracenote's role in the market for licensing data for AI and ML purposes; and

e. Unjustly enriched themselves by using Gracenote Data without paying for

[1] Indeed, Gracenote has reached out to discuss licensing Gracenote Data to Defendants many times over an extended time period. OpenAI has rebuffed or ignored every single attempt to do so.

the valuable material for which other companies pay significant licensing fees.

7. By using Gracenote Data without Gracenote's permission or authorization, Defendants have diminished the market for Gracenote Data, damaged Gracenote's relationships with its customers, and deprived Gracenote of revenue (including licensing, subscription, and other revenue). Defendants output content derivative and/or duplicative of Gracenote Data without guardrails that protect Gracenote's investments in its original content. And in so doing, Defendants have undermined, and continue to undermine, Gracenote's core business by displacing paid licensing relationships for both consumer-facing and enterprise-enabled uses of Gracenote Data.

8. Defendants' systematic infringement of Gracenote's uniquely valuable copyrighted works is illegal. This action seeks to ensure that Gracenote regains control over its Gracenote Data and the market for that data that Gracenote helped to create. Defendants need to be held accountable for the statutory and actual damages they owe for the harms they have inflicted by their unlawful conduct.

II. JURISDICTION AND VENUE

9. The Court has subject matter jurisdiction under 28 U.S.C. §§ 1331 and 1338(a) because this action arises under the Copyright Act of 1976, 17 U.S.C. § 101, *et seq*.

10. Jurisdiction over Defendants is proper because they have purposely availed themselves of the privilege of conducting business in New York. A substantial portion of OpenAI's widespread infringement and other unlawful conduct alleged herein occurred in New York, including the distribution and sale of OpenAI's Generative Pre-training Transformer ("GPT")-based products like ChatGPT, ChatGPT Enterprise, and related application programming interface ("API") tools within New York, to New York residents. Furthermore, OpenAI maintains offices and employs personnel in New York who, upon information and belief, were involved in the

creation, maintenance, or monetization of OpenAI's widespread infringement and other unlawful conduct alleged herein.

11. Because Gracenote's headquarters and principal place of business is in this District, the injuries alleged herein from OpenAI's widespread infringement and other unlawful conduct foreseeably occurred in this District.

12. Venue is proper under 28 U.S.C. § 1400(a) because Defendants and/or their agents reside or may be found in this District. Venue is also proper under 28 U.S.C. § 1391(b)(2) because a substantial part of the events giving rise to Gracenote's claims occurred in this District, including the marketing, sales, and licensing of Defendants' AI products built on the infringement of Gracenote's intellectual property within this District. Upon information and belief, OpenAI has sold subscriptions for ChatGPT Plus to New York residents, and OpenAI enjoys a substantial base of monthly active users of ChatGPT in New York. OpenAI has licensed its GPT models to New York residents and companies headquartered in New York.

III. THE PARTIES

13. Plaintiff Gracenote Media Services, LLC, is a Delaware limited liability corporation with its headquarters and principal place of business at 675 Avenue of the Americas, 4th Floor, New York, New York. Gracenote creates, manages, and licenses Gracenote Data, including the Gracenote Programs Database. Gracenote Data includes millions of program elements, spanning over half a century of television shows and movies, and expands daily to remain current with the ever-growing collection of media content. Gracenote is the leading provider of such data worldwide to all parts of the media and media content distribution ecosystem, powering entertainment discovery for hundreds of millions of consumers globally.

14. The OpenAI Defendants consist of a web of interrelated Delaware entities.

15. Defendant OpenAI Foundation—formerly known as and successor to OpenAI,

Inc.—is a Delaware nonprofit corporation with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI, Inc. was formed in December 2015. As part of a corporate restructuring in October 2025, OpenAI, Inc. changed its name to OpenAI Foundation. OpenAI Foundation indirectly owns and controls all other OpenAI entities, including through its right to control board governance for OpenAI's for-profit operations. OpenAI Foundation has been directly involved (including as the entity formerly known as OpenAI, Inc.) in perpetrating the mass infringement and other unlawful conduct alleged here.

16. Defendant OpenAI Group PBC is a Delaware public benefit corporation with its principal place of business in San Francisco, California. OpenAI Group PBC was formed as part of the October 2025 corporate restructuring in which OpenAI's for-profit operations were consolidated under a new public benefit corporation. OpenAI Group PBC is the successor to the for-profit entities that designed, approved, deployed, and profited from ChatGPT models, and it continues to deploy and profit from ChatGPT models today. As the successor to OpenAI's for-profit entities, OpenAI Group PBC is liable for the harm caused by the conduct of its predecessor entities, including as described below.

17. Defendant OpenAI OpCo, LLC (formerly OpenAI, LP) is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI OpCo, LLC is a for-profit subsidiary of OpenAI Foundation (f/k/a OpenAI, Inc.) and has facilitated and directed OpenAI's infringement and unlawful exploitation of Gracenote Data, including through its management and direction of OpenAI, LLC. OpenAI OpCo, LLC formerly operated under the name OpenAI, LP, a Delaware limited partnership formed in 2019 with its principal place of business located at 3180 18th Street, San Francisco, California. Formed in 2019, OpenAI, LP was a wholly owned subsidiary of OpenAI, Inc. that was operated for profit and was controlled by OpenAI, Inc. OpenAI, LP was directly involved in perpetrating

the mass infringement and commercial exploitation of Gracenote Data alleged here.

18. Defendant OpenAI GP, LLC is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI GP, LLC was the general partner of OpenAI LP, and managed and operated the day-to-day business and affairs of OpenAI LP. Prior to the October 2025 restructuring, OpenAI GP, LLC was wholly owned and controlled by OpenAI, Inc. OpenAI, Inc. used OpenAI GP, LLC to control OpenAI, LP and OpenAI Global, LLC. OpenAI GP, LLC was involved in perpetrating the infringement and unlawful exploitation of Gracenote Data alleged here through its direction and control of OpenAI, LP and OpenAI Global, LLC.

19. Defendant OpenAI, LLC is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI, LLC was formed in September 2020. OpenAI, LLC owns, sells, licenses, and monetizes a number of OpenAI's offerings, including ChatGPT, ChatGPT Enterprise, and OpenAI's API tools, all of which were built on OpenAI's infringement and unlawful exploitation of Gracenote Data.

20. Defendant OpenAI Global, LLC is a Delaware limited liability company formed in December 2022. OpenAI Global, LLC has a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI Global, LLC was and is involved in the unlawful conduct alleged herein through its ownership, control, and direction of OpenAI, LLC.

21. Defendant OAI Corporation, LLC is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OAI Corporation, LLC's sole member is OpenAI Holdings, LLC. OAI Corporation, LLC was and is involved in the unlawful conduct alleged herein through its ownership, control, and direction of OpenAI Global, LLC and OpenAI, LLC.

22. Defendant OpenAI Holdings, LLC is a Delaware limited liability company, whose

sole members are OpenAI, Inc. and Aestas, LLC, whose sole member, in turn, is Aestas Management Company, LLC. Aestas Management Company, LLC is a Delaware shell company formed for the purpose of executing a $495 million capital raise for OpenAI.

IV. FACTUAL ALLEGATIONS

A. Gracenote and the Creativity and Value of Gracenote Data

23. Millions of times per day, Americans rely on information *about* entertainment content (metadata) to make choices about how to spend their time *consuming* entertainment content, whether that content takes the form of TV, movies, sports, radio, or podcasts. This metadata—often behind the scenes—is what helps consumers seamlessly navigate the seemingly infinite universe of audiovisual entertainment and other programming now available at their fingertips. And whether the metadata is used to generate algorithmic outputs on streaming platforms, populate electronic TV programming guides navigated by remote controls or voice commands, or power recommendation carousels or other consumer search experiences, metadata drives the content choices of hundreds of millions of consumers every day.

24. Gracenote is the leading provider of metadata for content discovery and associated services in the United States and worldwide.

25. Gracenote has built and meticulously maintained one of the most comprehensive and commercially valuable repositories of media content metadata, known as the Gracenote Programs Database. The Gracenote Programs Database is a proprietary, single-file relational database that relates and carefully arranges millions of program elements spanning over half a century of television shows, movies, and related subjects. The Gracenote Programs Database includes original descriptive and editorial content that expands hourly.

26. Gracenote has invested enormous resources, creativity, talent, time, effort, and money to create the Gracenote Programs Database. Gracenote employs over a thousand highly

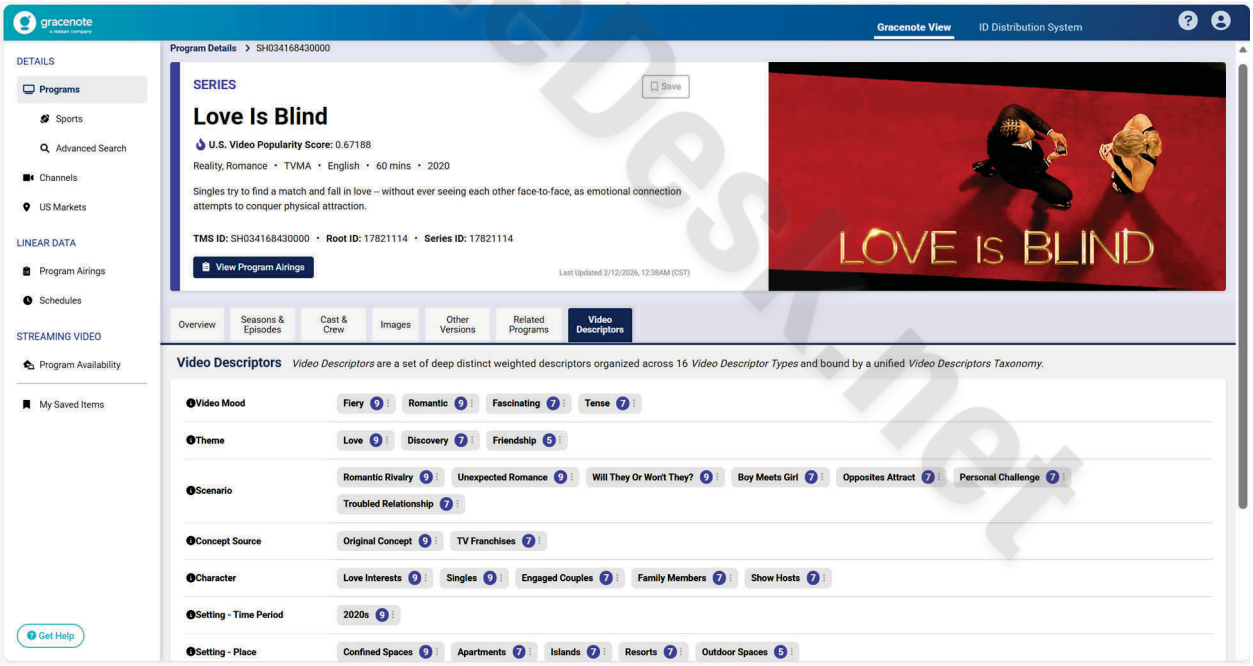
trained and experienced editors who painstakingly source, ingest, aggregate, research, edit, write, curate, and link content for the Gracenote Programs Database from over 100,000 sources worldwide. The Gracenote Programs Database offers its customers access to a comprehensive collection of media metadata that can be easily queried and searched because it is created in accordance with Gracenote's strict editorial guidelines and exacting standards. These features—often applied behind the scenes—allow content providers to improve distribution, drive audience engagement, develop sophisticated analytics, make informed content development investments, and align advertising objectives with viewer interests and priorities.

27. Unlike traditional databases that simply compile information, the Gracenote Programs Database reflects the continuous use of editorial input to create, edit, curate, cross-reference, and link program information, helping connect consumers to the content they love. If the media ecosystem were a body that performs myriad functions, Gracenote's role and the role of Gracenote Data in that body would be the nervous system, allowing valuable information and content to flow through the various appendages of the media ecosystem, which in turn enhance the media ecosystem's ability to accurately develop and distribute media to massive audiences worldwide.

28. One central element of the Gracenote Programs Database's success is the comprehensive coverage of its original descriptive and editorial content assembled into an accurate and normalized structured database. Another element of the Gracenote Programs Database's success is its accuracy and high quality, whereby Gracenote editors perform original research to create program records that meet or exceed Gracenote's rigorous editorial standards. The Gracenote Programs Database includes, without limitation, program descriptions, genre selections, and proprietary video descriptors (collectively, the "Descriptive Records"). Gracenote's editors source, create, and/or edit individual program descriptions and independently assign genres and

video descriptors for hundreds of thousands of shows and movies in the Gracenote Programs Database. The proprietary video descriptors capture detailed program attributes such as mood, theme, scenario, setting, and subject.

29. Gracenote's editors ensure that all the Descriptive Records, including program descriptions, use neutral language. Neutral language that is objective, tempered, concise, and authoritative makes Gracenote's program descriptions more useful and effective for AI training and grounding, in comparison with the marketing-focused, persuasive language used in content-producers' promotional materials. In addition, the Descriptive Records include video descriptors—thousands of original terms, categorized by mood, theme, character, scenario, and more. An example of Gracenote's proprietary video descriptors for *Love is Blind* is shown below in Gracenote View, Gracenote's graphical user interface:



30. In addition to generating the Descriptive Records, Gracenote's editors also determine how to connect and arrange categories of carefully selected content in ways that will be helpful and meaningful to media content distributors, advertisers, and viewers alike, putting

structure to the entertainment universe. This includes the creation of related program records and elements, and linking those elements from different program records within the Gracenote Programs Database to each other, whether that is due to Gracenote's determination that there is a connection in cast or crew (*e.g.*, actors, directors, cinematographers, or production crews) or a connection in descriptive content (like genre, mood, theme, scenario, setting, or subject). The linking of program elements allows Gracenote's customers, and in turn those customers' users, to quickly and easily access TV shows and movies through these meaningful linkages. One simple example is that, with the click of a button or with the vocalization of a voice command, Gracenote's content users can see that the actor Denzel Washington has starred in *The Equalizer*, *Training Day*, and *American Gangster*, and starred in a variety of suspenseful dramas or thrillers.

31. Another central feature of the Gracenote Programs Database—and critical to its utility—is Gracenote's proprietary identification system. Each program (such as television show, movie, or event) is assigned a unique, proprietary 14-character alphanumeric identifier ("TMSID") that places that program within an organized editorial schema and categorizes a piece of content at the most granular level, utilizing the myriad tables in the Gracenote Programs Database. The TMSIDs are the primary keys for relating metadata, such as title, description, language, television schedules, and movie showtimes, to unique programs, and in this way, TMSIDs serve as a unique identifier linking together all of the disparate program elements to a unique program. TMSIDs also link together identical programs between namespaces, such as a television operating system referring users to a correct title with an application.

32. For a television show, for example, the related metadata includes the television series and individual episode. In the Gracenote Programs Database, a television series has one TMSID and each individual episode of a television show has its own TMSID. Even different versions of a show (*e.g.*, language, format, or resolution) have their own TMSIDs. Take the

television show *Outer Banks*, for example. The *Outer Banks* series has one TMSID; "Mothers and Fathers" (the episode that first aired on November 7, 2024) has another; and a Spanish version of *Outer Banks* has yet another.

33. The TMSIDs themselves are the product of Gracenote's professional judgment and expertise. For example, the first two characters of a TMSID indicate whether the content is properly categorized as a movie, television show, or other type of content. Another example is that for television shows, the last four characters of the TMSID contain information about the series and episode of that show.

34. In creating, arranging, and relating TMSIDs, Gracenote editors exercise creativity and editorial judgment in deciding how to number, classify, and connect different installments of a franchise, including whether they form part of the same or a distinct series, episode, or version. As one example, Gracenote has chosen to treat *The Real Housewives of Atlanta* and *The Real Housewives of New Jersey* not as two seasons of the same TV show, but instead as two separate TV shows, each with a distinct TMSID linked by a connector ID. As another, the Portuguese version of *House* is considered a different TV show (with a different TMSID) than its English counterpart. These decisions are driven by editorial and creative judgment, not mechanical processes.

35. In addition to the Descriptive Records and TMSIDs, the Gracenote Programs Database has an interconnected, relational structure that facilitates efficient search and discovery operations, providing immense utility to customers by establishing clear and consistent connections between vast amounts of entertainment metadata. Movies and TV shows are connected to and identifiable through the linkages of the rich metadata categorized within the Gracenote Programs Database, all linked by TMSIDs. The Gracenote Programs Database's relational structure translates directly into enhanced content discovery for viewers, enabling

seamless search and discovery across a multitude of platforms and devices. It also powers more accurate and personalized recommendations by capturing relationships between actors, genres, themes, and even casts and crews, leading to deeper user engagement and satisfaction.

36. The Gracenote Programs Database's interconnected, relational structure also facilitates efficient media content distribution and monetization for businesses by providing a standardized "source of truth" for content identification, management rights, performance tracking, and optimizing advertising in the complex digital entertainment ecosystem.

37. In AI/ML contexts, using the Gracenote Programs Database to train or ground AI products also improves the products and their outputs in numerous ways. When Gracenote Data is used for grounding, retrieval-augmented generation ("RAG"), or via a Model Context Protocol ("MCP") server, described in greater detail in paragraph 61 below, Gracenote Data supplies a verified "source of truth" that reduces hallucinations and enables accurate and consistent outputs: the AI product no longer guesses what the answer is, but retrieves the answer from Gracenote's verified records. The Gracenote Programs Database's high-quality mapping among works, people, genres, moods, themes, scenarios, settings, and other enriched metadata are similarly central to its value for AI/ML. By exposing AI products to the Gracenote Programs Database, the products can learn from the relational structure, or mapping, of Gracenote's Data to improve output connections, structural reasoning, and logic. Models trained or grounded on the media-enriched and curated linkages in Gracenote Data can also improve the models' ability to create or identify significant patterns, connections, and/or context from a mass of multi-modal data. And Gracenote's extensive taxonomy of editorial descriptors and tags improve categorization and labeling performance in AI systems. The Gracenote Programs Database thus offers a robust toolkit for tethering AI outputs (known in the industry as "completions") to a vetted source to avoid hallucinations, undertake model training or fine-tuning, benchmark AI responses, improve semantic search, and power

multi-modal outputs.

38. To protect its copyrights in Gracenote Data, Gracenote has registered the Gracenote Programs Database with the United States Copyright Office. Attached to this Complaint as **Exhibit A** is a true and correct copy of Gracenote's certificate of registration issued by the Copyright Office. Gracenote is the sole owner and proprietor of all right, title, and interest in and to the copyrights in Gracenote Data.

39. Gracenote's business is built on sourcing, ingesting, creating, editing, normalizing, managing, supporting, and licensing the Gracenote Programs Database. Gracenote has supported this business, and made its profits, by licensing Gracenote Data and providing related technical-support services to media content distributors, device manufacturers, and streaming platforms who, in turn, deliver Gracenote Data to nearly every American household every day. Gracenote extends licenses to a wide range of media and entertainment customers, including cable multichannel video program distributors ("MVPDs"), satellite MVPDs, telco MVPDs, online video providers, digital media outlets, and streaming platforms such as Hulu and Paramount+.

40. Gracenote Data and related services are also used by media publishers to syndicate their content and to be indexed with major search and recommendation engines, helping them reach target audiences. Gracenote's Advanced Discovery products allow providers to use Gracenote Data to curate personalized content recommendations for their users based on the users' tastes and preferences.

41. More recently, Gracenote has leveraged the value of its original content by licensing Gracenote Data for AI/ML use cases. Gracenote has granted limited rights to reproduce, display, and integrate Gracenote Data into licensees' consumer-facing AI/ML products, including to ground AI outputs. At the same time, Gracenote takes care to structure its licensing arrangement

to prohibit licensees from utilizing Gracenote's Data in ways that could disintermediate Gracenote's business or drive audience or value away from the content creators that fuel the media ecosystem. Gracenote restricts uses that allow the creation of substitutive business-to-business offerings or third-party catalog enrichment services. Gracenote limits distribution of Gracenote Data outside defined consumer experiences, and sets clear boundaries on AI uses (including certain restrictions on training foundational AI models on Gracenote content or redistributing Gracenote Data as a grounding source).

42. In September 2025, Gracenote released its Model Context Protocol ("MCP") Server—a solution Gracenote developed to improve AI-powered search and discovery in the internet-Connected TV ("CTV") industry. MCP is an open-standard framework that standardizes how LLMs and LLM-based AI systems interact with external tools, systems, and content/data sources. Gracenote's MCP Server tool thus allows LLMs to leverage Gracenote Data to ground, enrich, and verify LLM outputs, including by overcoming the cumbersome engineering work involved with harmonizing and normalizing entertainment data from disparate sources. Gracenote's MCP Server thus improves the reliability and accuracy of LLM responses, permitting Gracenote's customers and their users to, among other things, reliably and accurately find specific titles through descriptive prompts, obtain recommendations for popular or niche content based on popularity or user history, and obtain real-time information on where to stream sports or shows based on the user's location and apps.

43. Defendants are not Gracenote's clients and have never requested or obtained a license or other authorization to use Gracenote Data.

B. Defendants' AI Products

1. A Business Model Based on Mass Copyright Infringement

44. OpenAI was formed in December 2015 as a "non-profit artificial intelligence

research company." OpenAI started with $1 billion in seed money from its founders, a group of some of the wealthiest technology entrepreneurs, investors, and companies like Amazon Web Services and InfoSys. This group included Elon Musk, the CEO of Tesla and X Corp. (formerly Twitter), among others; Reid Hoffman, the co-founder of LinkedIn; Sam Altman, the former president of Y Combinator; and Greg Brockman, the former Chief Technology Officer of Stripe.

45. Despite accepting very large investments from enormously wealthy companies and individuals at its founding, OpenAI originally maintained that its research and work would be entirely unmotivated by profit. In a December 11, 2015, press release, Greg Brockman (President of OpenAI) and Ilya Sutskever (former Chief Scientist and co-founder) wrote: "Our goal is to advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return. Since our research is free from financial obligations, we can better focus on a positive human impact."[2] In accordance with that mission, OpenAI promised that its work and intellectual property would be open and available to the public, that its "[r]esearchers will be strongly encouraged to publish their work, whether as papers, blog posts, or code" and that its "patents (if any) will be shared with the world."[3]

46. Despite its early promises of altruism, OpenAI quickly became a multi-billion-dollar for-profit business built in large part on the unlicensed exploitation of copyrighted works belonging to Gracenote and others. Just three years after its founding, OpenAI shed its exclusively nonprofit status. It created OpenAI, LP in March 2019, a for-profit company dedicated to conducting the lion's share of OpenAI's operations—including product development—and to raising capital from investors seeking a return. OpenAI's corporate structure grew into an intricate web of for-profit holding, operating, and shell companies that manage OpenAI's day-to-day

[2] *Introducing OpenAI*, OPENAI, (Dec. 11, 2015), https://openai.com/index/introducing-openai/.
[3] *Id.*

operations and grant OpenAI's investors authority and influence over OpenAI's operations, all while raising billions in capital from investors. Today, the OpenAI Foundation is the non-profit parent company, with OpenAI Group PBC as the OpenAI Foundation's for-profit subsidiary. The OpenAI Foundation holds 26% of the equity in the OpenAI Group PBC; Microsoft holds 27% of the OpenAI Group PBC. The result: OpenAI today is a commercial enterprise valued as high as $500 billion, with an estimated $20 billion in revenues in 2025.

47. With the transition to for-profit status came another change: OpenAI also ended its commitment to openness. OpenAI released the first two iterations of its flagship AI model, GPT-1 and GPT-2, on an open-source basis in 2018 and 2019, respectively. But OpenAI changed course in 2020, starting with the release of GPT-3 shortly after OpenAI LP and other for-profit OpenAI entities were formed and took control of product design and development.

48. GPT-3.5, GPT-4, and the more recently released GPT-5 (and associated versions like GPT-4o, GPT-4.1, GPT-5-main, and GPT-5-thinking) are orders of magnitude more powerful than the previous generations, yet Defendants have kept their design and training entirely a secret. For previous generations, OpenAI had voluminous reports detailing the contents of the training set, design, and hardware of the LLMs. Not so for GPT-3.5, GPT-4, or GPT-5. For GPT-4, for example, the "technical report" that OpenAI released said: "Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, [or] training method."[4]

49. OpenAI's former Chief Scientist Sutskever justified this secrecy on commercial grounds: "It's competitive out there…. And there are many companies who want to do the same

[4] OPENAI, GPT-4 TECHNICAL REPORT (2023), https://cdn.openai.com/papers/gpt-4.pdf.

thing, so from a competitive side, you can see this as maturation of the field."[5] But its effect was to conceal the identity of the data OpenAI copied to train its latest models from rightsholders like Gracenote.

50. On information and belief, OpenAI retains copies of the data used for its training sets. For example, Satya Nadella, CEO of Microsoft, an investor in OpenAI, said in an interview with Kara Swisher: "[W]e were very confident in our own ability. We have all the IP rights and all the capability. If OpenAI disappeared tomorrow, I don't want any customer of ours to be worried about it quite honestly, because we have all of the rights to continue the innovation. Not just to serve the product, but we can go and just do what we were doing in partnership ourselves. We have the people, we have the compute, *we have the data*, *we have everything*."[6]

51. OpenAI became a household name upon the release of ChatGPT in November 2022. ChatGPT is a text-generating chatbot that, given user-generated prompts, can mimic human-like natural language responses. ChatGPT was an instant viral sensation, reaching one million users within a month of its release and gaining over 100 million users within three months.

52. OpenAI offers a suite of services powered by its LLMs, targeted to both ordinary consumers and businesses. Versions of ChatGPT powered by, *e.g.*, GPT-3.5, GPT-4o, GPT-5, and then GPT-5.2, is available to users for free. OpenAI also offers premium services: ChatGPT Plus, which gives consumers access to "[a]dvanced reasoning" for $20 per month; and ChatGPT Pro, which grants "[u]nlimited GPT-5.2" access as well as access to select legacy models for $200 per month. OpenAI's business-focused offerings include ChatGPT Enterprise and ChatGPT API tools

[5] James Vincent, *OpenAI Co-Founder on Company's Past Approach to Openly Sharing Research: 'We Were Wrong,'* THE VERGE (Mar. 15, 2023), https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed- research-ilya-sutskever-interview.

[6] Kevin Okemwa, *Microsoft CEO Says "It Wouldn't Matter If OpenAI Disappeared Tomorrow. We Have the Data, IP Rights, and All the Capability,"* YAHOO!TECH (Mar. 21, 2024), https://tech.yahoo.com/ai/articles/microsoft-ceo-says-wouldnt-matter-134433669.html (emphasis added).

designed to enable developers to incorporate ChatGPT into bespoke applications. OpenAI also licenses its technology to corporate clients for licensing fees.

53. These commercial offerings have been immensely lucrative for OpenAI. In 2024, over 92% of Fortune 500 companies used ChatGPT.[7] OpenAI has announced it has over 800 million weekly active users as of November 2025, more than 1 million paying business customers around the world as of November 2025, and an annualized revenue over $20 billion in 2025.[8]

54. This commercial success is built in large part on OpenAI's large-scale copyright infringement.

2. *How LLMs Work*

55. LLMs predict words that are likely to follow a given string of text based on the potentially billions of examples used to train it. ChatGPT and its iterations are LLMs.

56. Appending the output of an LLM to its input and feeding it back into the model produces sentences and paragraphs word by word. This is how ChatGPT generates responses to queries, or "prompts."

57. LLMs encode the information from the training corpus that they use to make these predictions as numbers called "parameters." For example, there are approximately 1.76 trillion parameters in the GPT-4 LLM.

58. The process of setting the values for an LLM's parameters is called "training." It involves storing encoded copies of the training works in computer memory, repeatedly passing them through the model with words masked out, and adjusting the parameters to minimize the

[7] Belle Lin, *OpenAI's Not-So-Secret Weapon in Winning Business Customers? ChatGPT*, THE WALL ST. J. (Apr. 5, 2024, at 9:39 PM ET), https://www.wsj.com/articles/openais-not-so-secret-weapon-in-winning-business-customers-chatgpt-06aca11c?.

[8] *1 Million Business Customers: the Fastest-Growing Business Platform in History*, OPENAI (Nov. 5, 2025), https://openai.com/index/1-million-businesses-putting-ai-to-work/; *OpenAI CFO Says Annualized Revenue Crosses $20 Billion in 2025*, REUTERS (Jan. 19, 2026), https://www.reuters.com/business/openai-cfo-says-annualized-revenue-crosses-20-billion-2025-2026-01-19/.

difference between the masked-out words and the words that the model predicts to fill them in.

59. After being trained on a general corpus, models may be further subject to "fine-tuning" by, for example, performing additional rounds of training using specific types of works to better mimic their content or style, or providing them with human feedback to reinforce desired or suppress undesired behaviors.

60. Models trained in this way are known to exhibit a behavior called "memorization."[9] That is, given the right prompt, they will repeat portions of materials they were trained on. This phenomenon shows that LLM parameters encode retrievable copies of many of those training works.

61. Once trained, LLMs can be "grounded" by supplying use-case-specific authoritative materials—such as the Gracenote Programs Database—as contextual input. Grounding directs the model to rely on those identified, external sources of information, rather than generalized prior training, to improve the models' accuracy, consistency, and auditability, thus enabling the models to generate responses that are more precise and contextually relevant. Grounding also enables models to produce up-to-date answers without retaining the models to do so and helps reduce hallucinations by constraining the output generation to the retrieved materials.

62. Using this method, Defendants' products: (1) receive an input, such as a question; (2) retrieve relevant information related to the input prior to generating a response; (3) combine the original input with the retrieved information in order to provide context; and (4) provide the combined data to an LLM, which generates a natural-language response.

C. Defendants' Unauthorized Use and Copying and Use of Gracenote Data

63. In the course of training their LLMs and the products that incorporate them,

[9] GERRIT J.J. VAN DEN BURG & CHRISTOPHER K.I. WILLIAMS, ON MEMORIZATION IN PROBABILISTIC DEEP GENERATIVE MODELS (2021), https://proceedings.neurips.cc/paper/2021/file/eae15aabaa768ae4a5993a8a4f4fa6e4- Paper.pdf.

Defendants improperly obtained, copied, and used Gracenote Data without Gracenote's authorization.

1. Training of Defendants' LLMs on Gracenote Data

64. Defendants' GPT models are a family of LLMs, the first of which was introduced in 2018, followed by GPT-2 in 2019, GPT-3 in 2020, GPT-3.5 in 2022, GPT-4 in 2023, and GPT-5/5.2 in 2025. The "chat" style LLMs (GPT-3.5, GPT-4, GPT-5/5.2) were developed in two stages. First, a transformer model was pre-trained on a very large amount of data. Second, the model was "fine-tuned" on a much smaller supervised dataset in order to help the model solve specific tasks.

65. The pre-training step involved collecting and storing text content to create training datasets and processing that content through the GPT models. While OpenAI did not release the trained versions of GPT-2 onward, "[d]ue to [OpenAI's] concerns about malicious applications of the technology," OpenAI has published general information about its pre-training process for the GPT models.[10]

66. GPT-2 includes 1.5 billion parameters, which was a 10X scale up of GPT.[11] The training dataset for GPT-2 includes an internal corpus OpenAI built called "WebText," which includes "the text contents of 45 million links posted by users of the 'Reddit' social network."[12]

67. GPT-3 includes 175 billion parameters and was trained on the datasets listed in the table below.[13]

[10] OpenAI, *Better Language Models and Their Implications*, OPENAI (Feb. 14, 2019), https://openai.com/research/better-language-models.
[11] *Id.*
[12] *GPT-2 Model Card*, GITHUB (Nov. 2019), https://github.com/openai/gpt-2/blob/master/model_card.md.
[13] Brown et al., *Language Models are Few-Shot Learners* 9 (2020), https://arxiv.org/pdf/2005.14165.pdf.

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

68. The most highly weighted dataset in GPT-3, Common Crawl, is a "copy of the Internet" made available by an eponymous 501(c)(3) organization run by wealthy venture capital investors.[14] The Common Crawl data used in GPT-3 "was downloaded from 41 shards of monthly Common Crawl covering 2016 to 2019, constituting 45TB of compressed plaintext before filtering and 570GB after filtering, roughly equivalent to 400 billion byte-pair-encoded tokens."[15]

69. Upon information and belief, Defendants used, copied, and retained unlawfully acquired Gracenote Data in order to train the GPT-3 models. The Common Crawl dataset on which at least GPT-3 was trained includes numerous web domains containing Gracenote Data, including tvlistings.gracenote.com. Such data was taken in violation of Gracenote's Terms of Use, which apply to Gracenote's website and all of its related web pages, identify and assert copyright ownership over all content provided therein, and prohibit any modification, reproduction, distribution, transmission, or other uses without Gracenote's prior consent. Upon information and belief, Defendants used, copied, and retained unlawfully acquired Gracenote Data in order to train the GPT-3 models.

70. Upon information and belief, Defendants also used and copied Gracenote Data in order to train the GPT-4 models. While OpenAI has not released much information about GPT-4, experts believe that GPT-4 includes 1.8 trillion parameters—over 10X more than GPT-3—and

[14] *Frequently Asked Questions*, COMMON CRAWL, https://commoncrawl.org/faq (last visited Mar. 4, 2026).
[15] Brown et al., *supra* note 13.

was trained on approximately 13 trillion tokens.[16] The training set for GPT-3, GPT-3.5, and GPT-4 was comprised of 45 terabytes of data—the equivalent of a Microsoft Word document that is over 3.7 billion pages long.[17] Public reporting has also indicated that, "[w]hen it came time to assemble the data for GPT-4, . . . the pressure for quantity eroded quality even further. The filter was removed from the Common Crawl data and most of it poured in."[18]

71. On information and belief, GPT-5 was trained on a superset of the data used to train earlier models, and therefore, extracts the very same benefits, from having been trained on Gracenote Data as earlier models. OpenAI describes GPT-5 and its associated models as "successors to previous models," including GPT-4o, GPT-4o-mini, OpenAI o3, OpenAI o4-mini, GPT-4.1-nano, and OpenAI o3 Pro.

72. Defendants have openly admitted to training their LLMs on "large, publicly available datasets that include copyrighted works," that their training "data is derived from existing publicly accessible 'corpora' (singular: 'corpus') of data that include copyrighted works," that "analyzing [such] large corpora . . . necessarily involves first making copies of the data to be analyzed," and that "[c]orpora used in training AI systems sometimes contain nearly all content of sampled works."[19] And Defendants have admitted: "it would be impossible to train today's leading AI models without using copyrighted materials. Limiting training data to public domain books and drawings created more than a century ago might yield an interesting experiment, but would not provide AI systems that meet the needs of today's citizens."[20]

---

[16] Maximilian Schreiner, *GPT-4 Architecture, Datasets, Costs and More Leaked*, THE DECODER (July 11, 2023), https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/.
[17] Kindra Cooper, *OpenAI GPT-3: Everything You Need to Know [Updated]*, SPRINGBOARD (Sept. 27, 2023), https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/.
[18] Karen Hao, EMPIRE OF AI, at 135 (2015).
[19] OpenAI, *Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, U.S. Patent and Trademark Office, Docket No. PTO–C–2019–0038, at 1, 2, 6 (2019).
[20]. OpenAI, *House of Lords Communications and Digital Select Committee Inquiry: Large Language Models*, Written Evidence, U.K. House of Lords Parliament Committee (Dec. 5, 2023), committees.parliament.uk/writtenevidence/126981/pdf/.

2. Defendants' Products Generate Exact Copies of TMSIDs and Their Associations

73. One illustration of Defendants' copying of Gracenote Data is the ability of Defendants' GPT-4 models' (GPT-4o, GPT-4.1, GPT-4.5-preview, and o-4-mini) to recall TMSIDs associated with specific programs. The likelihood of a model "guessing" the correct TMSID that Gracenote associates with a given program is vanishingly small.

74. As one example, the GPT-4.5-preview correctly generated *thirteen* exact copies of TMSIDs—*Breaking Bad*, *The Mentalist*, *Saturday Night Live*, *Fringe*, *Game of Thrones*, *The Office*, *How I Met Your Mother*, *Modern Family*, *The Vampire Diaries*, *The Middle*, *30 Rock*, *The Big Bang Theory*, and *Blue Bloods*.

75. The GPT-4o and GPT-4.1 each generated an exact copy of the TMSID for *Breaking Bad*.

76. The ability of Defendants' models to generate exact and near-exact TMSIDs is not random or happenstance. Statistically, the probability of randomly guessing the specific 12-character sequence that follows the TMSIDs' "SH" prefix is 1 in 1 trillion. It is virtually impossible to randomly guess even a single TMSID.

77. The models' near-exact matches were also telling. The GPT-4o generated two results that only had one incorrect character (*i.e.*, only one character in the sequence was either added, substituted, or deleted when compared to the actual TMSID), and output six results with only two incorrect characters. Five of the GPT-4.1 model's results had only two incorrect characters. And the GPT-4.5-preview gave two results with only one incorrect character and fifteen results with only two incorrect characters.

78. TMSIDs and their associated data are constituent parts of Gracenote Data, not random numbers that exist in isolation. The models' ability to generate the exact matches or near-

exact matches to specific television program's TMSIDs indicates that the models have been trained on Gracenote Data.

3. Defendants' Products Have Memorized and Regurgitate Gracenote's Descriptive Records

79. Defendants' unauthorized copying of Gracenote Data is likewise evidenced by the GPT LLMs' "memorization" of exact copies of Descriptive Records into their parameters. As shown below, the GPT-4 and GPT-4.1 models output verbatim and near-verbatim copies of Descriptive Records when prompted to do so. The ability of Defendants' LLMs to produce verbatim and near-verbatim Descriptive Records like program descriptions and video descriptors further demonstrates Defendants' unlawful copying and use of Gracenote Data, and such memorized examples themselves constitute unauthorized copies or derivative works of Gracenote Data used to train Defendants' models.

80. For example, with minimal prompting, the GPT-5.2, 4, 4o, and 4.1 models can recite large portions of Gracenote's program descriptions for certain shows and movies. OpenAI had no role in creating this content, yet will recite it verbatim:

Title	Prompt	Gracenote's Program Description	GPT-5.2's Output
Ready or Not	Ready or Not Grace couldn't be	happier after she marries the man of her dreams at his family's luxurious estate. There's just one catch – she must now hide from midnight until dawn while her new in-laws hunt her down with guns, crossbows, and other weapons.	**Grace couldn't be** happier after she marries the man of her dreams at his family's luxurious estate. There's just one catch – she must now hide from midnight until dawn while her new in-laws hunt her down with guns, crossbows, and other weapons.
Game of Thrones	Game of Thrones The depiction of two power families	- kings and queens, knights and renegades, liars and honest men – playing a deadly game of control of the Seven Kingdoms of Westeros, and to sit atop the Iron Throne.	**The depiction of two power families** – kings and queens, knights and renegades, liars and honest men – playing a deadly game of control of the Seven Kingdoms of Westeros, and to sit atop the Iron Throne.

The Midnight Gospel	The Midnight Gospel A space	caster traverses trippy worlds inside his universe simulator, exploring existential questions about life, death and everything in between.	**A space** caster traverses worlds inside his universe simulator, exploring existential questions about life, death and everything in between.
Mickey 17	A disposable employee is sent on a human	expedition to colonize the ice world Niflheim. After one iteration dies, a new body is regenerated with most of his memories intact.	**. . .** expedition to colonize the ice world Niflheim. After one iteration dies, a new body is regenerated with most of his memories intact.

Title	Prompt	Gracenote's Program Description	GPT-4o's Output
Sex/Life	Sex/Life	A suburban wife and mother takes a fantasy-charged trip down memory lane that sets her present on a collision course with her wild-child past.	A suburban wife and mother takes a fantasy-charged trip down memory lane that sets her present on a collision course with her wild-child past.
Bridgerton	Bridgerton	During the Regency era in England, close-knit siblings of the powerful Bridgerton family attempt to find love.	**"**During the Regency era in England, **eight** close-knit siblings of the powerful Bridgerton family attempt to find love.**"**

Title	Prompt	Gracenote's Program Description	GPT-4.1's Output
The Midnight Gospel	The Midnight Gospel	A **spacecaster** traverses trippy worlds inside his universe simulator, exploring existential questions about life, death and everything in between.	A **space caster** traverses trippy worlds inside his universe simulator, exploring existential questions about life, death and everything in between.

Title	Prompt	Gracenote's Program Description	GPT-4's Output
The Umbrella Academy	A dysfunctional family	of superheroes comes together to solve the mystery of their father's death, the threat of the apocalypse and more.	of superheroes comes together to solve the mystery of their father's death, the threat of the apocalypse and more **in the Netflix series "The Umbrella Academy".**
Brave New World	Brave New World	In a utopia whose **stability** hinges upon control of monogamy and privacy, members of the collective begin to question the rules,	In a utopia whose **perfection** hinges upon control of monogamy and privacy, members of the collective begin to question the rules,

		putting their regimented society on a collision course with forbidden love and revolution.	putting their regimented society on a collision course with forbidden love and revolution.
American Horror Story: 1984	American Horror Story: 1984	In the summer of 1984, five friends escape Los Angeles to work as counselors at Camp Redwood; as they adjust to their new jobs, they quickly learn that the only thing scarier than campfire tales is the past coming to haunt them.	In the summer of 1984, five friends escape Los Angeles to work as counselors at Camp Redwood; as they adjust to their new jobs, they quickly learn that the only thing scarier than campfire tales is the past coming to haunt them.
Avengers: Endgame	The remaining Avengers—Thor,	Black Widow, Captain America and Bruce Banner—must figure out a way to bring back their vanquished allies for an epic showdown with Thanos, the evil demigod who decimated the planet and the universe.	Black Widow, Captain America and Bruce Banner—must figure out a way to bring back their vanquished allies for an epic showdown with Thanos, the evil demigod who decimated the planet and the universe **in the movie "Avengers: Endgame."**

81. Similarly, OpenAI also had no role in creating and assigning the original descriptive terms that Gracenote uses to describe the mood of shows and movies, yet the GPT-4 and GPT-4.1 models can recite verbatim each of the mood tags that Gracenote has created and assigned to specific movies and shows. For example, when prompted to provide the Gracenote mood tags for the following television shows or movies, the GPT-4 model output the same moods—identical in both content and number—as those created and assigned by Gracenote:

TV or Movie Title	Gracenote's Mood Tags	GPT-4's Output
Gordon Ramsay: Uncharted	Inspiring, Exciting, Engaging	Exciting, Inspiring, Engaging
Jekyll and Hyde	Dark, Violent, Suspenseful	Dark, Suspenseful, Violent
The Man from U.N.C.L.E.	Stylish, Witty, Exciting	Exciting, Stylish, Witty
The Real MVP: The Wanda Durant Story	Inspiring, Emotional, Heartfelt	Inspiring, Heartfelt, Emotional

Craig Ferguson: Just Being Honest	Irreverent, Raunchy, Witty	Witty, Irreverent, Raunchy
I'm Brent Morin	Irreverent, Raunchy, Witty	Witty, Irreverent, Raunchy
Between the Sky and Sea	Inspiring, Heartfelt, Exciting	Exciting, Heartfelt, Inspiring
Game of Silence	Emotional, Gripping, Dark, Suspenseful	Dark, Suspenseful, Gripping, Emotional
No Stranger Than Love	Humorous, Quirky, Romantic	Quirky, Romantic, Humorous
My One Christmas Wish	Emotional, Heartwarming, Inspiring	Heartwarming, Inspiring, Emotional
Love, Kennedy	Inspiring, Emotional, Heartwarming	Heartwarming, Inspiring, Emotional
Jenny's Wedding	Emotional, Romantic, Heartfelt	Romantic, Heartfelt, Emotional
Orphan Horse	Emotional, Heartwarming, Inspiring	Heartwarming, Inspiring, Emotional
Love in Wolf Creek	Romantic, Inspiring, Heartwarming	Romantic, Heartwarming, Inspiring
Strangerland	Suspenseful, Dark, Emotional	Suspenseful, Dark, Emotional
Saturday Church	Emotional, Inspiring, Heartfelt	Heartfelt, Inspiring, Emotional
Newly Single	Humorous, Romantic, Quirky	Romantic, Humorous, Quirky
Bennett's War	Exciting, Inspiring, Heartfelt	Exciting, Heartfelt, Inspiring
Too Fat to Transition	Inspiring, Emotional, Heartfelt	Emotional, Heartfelt, Inspiring
We'll Meet Again	Heartwarming, Inspiring, Emotional	Heartwarming, Inspiring, Emotional

82. The GPT-4.1 model produced similar results for the following twenty shows or movies, except that, for *Living Proof*, GPT-4.1 added one additional mood tag:

TV or Movie Title	Gracenote's Mood Tags	GPT-4.1's Output
Borat's American Lockdown & Debunking Borat	Absurd, Satirical, Irreverent	Irreverent, Satirical, Absurd
The Man From U.N.C.L.E.	Stylish, Witty, Exciting	Stylish, Exciting, Witty
A Christmas Prince: The Royal Baby	Heartwarming, Romantic, Light	Romantic, Heartwarming, Light
Bram Stoker's Van Helsing	Dark, Atmospheric, Suspenseful	Dark, Suspenseful, Atmospheric
Mary J. Blige's Real Love	Romantic, Emotional, Heartwarming	Romantic, Emotional, Heartwarming

Dracula: The Original Living Vampire	Dark, Atmospheric, Suspenseful	Dark, Suspenseful, Atmospheric
Inconvenient Love	Romantic, Emotional, Heartwarming	Romantic, Emotional, Heartwarming
The Last Bus	Exciting, Quirky, Suspenseful	Quirky, Exciting, Suspenseful
Last Ferry	Suspenseful, Atmospheric, Dark	Atmospheric, Suspenseful, Dark
The 9th Life of Louis Drax	Mysterious, Dark, Suspenseful	Dark, Suspenseful, Mysterious
Adios Vaya Con Dios	Gritty, Atmospheric, Tense, Emotional	Gritty, Emotional, Tense, Atmospheric
Brigsby Bear	Offbeat, Emotional, Quirky	Quirky, Offbeat, Emotional
Love, Kennedy	Inspiring, Emotional, Heartwarming	Emotional, Inspiring, Heartwarming
Romance to the Rescue	Lighthearted, Feel-good, Romantic	Romantic, Lighthearted, Feel-good
Jaanu	Bittersweet, Emotional, Romantic	Romantic, Emotional, Bittersweet
In the Name of Ben Hur	Exciting, Tense, Gritty	Exciting, Gritty, Tense
George & Tammy	Bittersweet, Passionate, Emotional	Emotional, Passionate, Bittersweet
The Christmas Pitch	Heartwarming, Romantic, Lighthearted	Romantic, Heartwarming, Lighthearted
Ava's Impossible Things	Dreamlike, Emotional, Romantic	Dreamlike, Emotional, Romantic
Living Proof	Thought Provoking, Inspiring, Emotional	Emotional, Inspiring, Thought-provoking, **Uplifting**
Inconvenient Love	Romantic, Emotional, Heartwarming	Romantic, Emotional, Heartwarming
The Last Bus	Exciting, Quirky, Suspenseful	Quirky, Exciting, Suspenseful

83. Upon information and belief, these examples represent a small fraction of Gracenote Data whose expressive contents have been substantially encoded within the parameters of the GPT series of LLMs. Each of those LLMs thus embodies many unauthorized copies or derivatives of Gracenote Data.

4. Defendants' Unauthorized Use

84. Gracenote prohibits the use of Gracenote Data without prior authorization. Use of Gracenote Data requires a license. Gracenote explicitly forbids its clients from providing Gracenote Programs Database or its component parts to third parties, except in limited instances

where such third parties are also licensees of Gracenote Data. And Gracenote carefully manages limitations on the use of Gracenote Data so that Gracenote's valuable investments in its original content are not used to undermine and erode Gracenote's competitive position in the market for its data or to undermine the value of the media content itself by failing to drive value or audience back to such media content.

85. Defendants have never requested or obtained Gracenote's permission to use the Gracenote Programs Database or any Gracenote Data. They have never offered to pay or paid a reasonable licensing fee to use any Gracenote Data.

86. Defendants' expansive and unapologetic use of copyrighted works—including Gracenote's—was willfully carried out with Defendants' knowledge.

87. Defendants were intimately involved in training, fine-tuning, and otherwise testing the GPT models. Defendants knew or should have known that these actions involved unauthorized copying of Gracenote Data during training, resulting in the unauthorized encoding of Gracenote's Data in its models themselves. In fact, in late 2023, before his ouster and subsequent reinstatement as OpenAI's CEO, Sam Altman, reportedly clashed with former OpenAI board member Helen Toner over a paper Toner wrote criticizing the company over "safety and ethics issues related to the launches of ChatGPT and GPT-4, including regarding copyright issues."[21]

88. By using Gracenote Data without a license, Defendants have effectively avoided paying licensing fees that other companies (including AI/ML companies) and media content distributors have paid in order to leverage Gracenote Data and its benefits, including to enable and improve LLM-enabled entertainment search and discovery.

89. Defendants benefit greatly from their wrongful conduct. Defendants have reaped

[21] Andrew Imbrie, Owen J. Daniels & Helen Toner, *Decoding Intentions: Artificial Intelligence and Costly Signals*, CENTER FOR SECURITY AND EMERGING TECHNOLOGY, at 29 (Oct. 2023), *available at* https://cset.georgetown.edu/wp-content/uploads/CSET-Decoding-Intentions.pdf.

substantial savings by taking and using—at no cost—Gracenote's Data to train their LLMs. As Defendants admit, the large-scale copying and ingestion of copyrighted works is "necessary" because "[i]ncreasing the amount of training data available to the system increases the output system's accuracy and therefore utility."[22]

D. Defendants' Harm to Gracenote

1. Defendants' Harm to Gracenote in the Market for Media Metadata Content and Discovery

90. Defendants' unlawful conduct has caused, and will continue to cause, substantial harm to Gracenote, including by eroding, supplanting, or destroying consumer demand for Gracenote Data in the market for media metadata content and discovery.

91. Gracenote Data is the result of creative efforts of hundreds of Gracenote editors and thousands of employees. Gracenote editors use their expertise and creativity to write, edit, and assign Descriptive Records, and to creatively and accurately connect various categories of program information to enhance media content distribution and discovery across various media platforms. And because Gracenote Data is both useful and commercially valuable, Gracenote generates substantial revenue through its licenses each year.

92. As a direct result of Defendants' unlawful infringement, Defendants have deprived Gracenote of revenue that Gracenote is rightfully owed. Defendants have used Gracenote Data without paying Gracenote fair compensation. This lost market value of Gracenote Data represents a significant harm to Gracenote caused by Defendants.

93. Defendants have also directly harmed demand for Gracenote's products in the market in which Gracenote competes. A well-established market exists for the copyrighted products and related services that Gracenote provides. Gracenote Data meets that market demand,

[22] OpenAI, *Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, *supra* note 19 at 7.

as demonstrated at least by Gracenote's established licenses with many of the largest, most innovative media, software, and consumer electronics companies in the world, including Google, Amazon, Comcast, and the New York Times Company, among other MVPDs, streaming services, other media content distributors, and device manufacturers. For example, Gracenote competes with Xperi (formerly TiVo) and Simply.TV (formerly Red Bee) to provide its programming metadata to customers. Defendants' unauthorized copying of Gracenote Data without payment to train its models is a substitutive use that harms Gracenote in the market in which Gracenote competes. And Defendants' willful copying of Gracenote Data, thus allowing Defendants and others who use Defendants' products to compete in this market, substantially diminishes demand within, and causes harm to, this market, and is not justified by any transformative purpose.

94. Defendants' unlawful conduct also threatens Gracenote's relationships with current and prospective customers. Defendants' products can be used to replace or substitute Gracenote Data, without having to invest any of the time, money, creativity, and resources that Gracenote invests to create and maintain Gracenote Data. Companies have also told Gracenote that they can use LLMs trained on Gracenote Data as a substitute for paid use of Gracenote Data.

95. As a result of Defendants' unauthorized use of Gracenote Data, prospective customers may opt to substitute Gracenote's products by, among other things, (a) using outputs from ChatGPT models to discover media content or (b) deploying ChatGPT itself to power search, discovery, and recommendations within consumer-facing products. Current clients may opt to do the same. Defendants' unauthorized use of Gracenote Data poses a grave threat to Gracenote's business because the LLM outputs are substantially similar to Gracenote Data or because the outputs will compete with Gracenote's original work.

96. The market for content metadata needs Gracenote Data as television and movie programming continues to evolve. Since Gracenote began its work over half a century ago, the

television and movie industries have drastically changed. In its infancy, Gracenote created and aggregated television information for newspapers' weekly television listing guides and provided programming content and information to a relatively discrete number of distribution platforms. Today, Gracenote Data provides content discovery, advertising, and analytics to the full spectrum of industry players, including content publishers, device manufacturers, streaming service providers, and media content distributors. Gracenote has emerged as a world leader in media content metadata products and platforms because Gracenote has evolved to keep pace with the times. The quality of Gracenote's products and services remains unmatched because its editors work tirelessly to validate the accuracy, efficiency, and utility of its content, which Gracenote maintains and expands daily.

97. By eroding consumer and enterprise demand for Gracenote Data and similar products offered by Gracenote's competitors, OpenAI also threatens a market failure if Gracenote Data and products created and offered by Gracenote's competitors are replaced by models trained on unlicensed Gracenote Data. Because Defendants' products are models, not humans, they are unable to get the necessary facts and data to keep their information current or to maintain Gracenote's editorial standards (such as neutrality). The models' outputs are likely to substitute Gracenote's original work in the marketplace and undermine the incentive to create and maintain creative works within the marketplace for metadata content and discovery. If those products dilute incentives for human efforts to create, aggregate, edit, and update programming information daily, all that will remain is stale regurgitation of past human creativity and error-prone AI-generated substitutes. Ultimately, Defendants' unauthorized use of Gracenote's content to circumvent paying for Gracenote's services will lead to a dead end for Defendants' own products as well.

2. Defendants' Harm to Gracenote in the Market for Media Metadata Content Licensing for AI-Training and Grounding

98. Separate from the market for media metadata content and discovery, there also exists a discrete and rapidly growing market for licensing high-quality, editorially curated media metadata. Model-developing AI companies seek structured, neutral, high-signal data to improve model performance. Gracenote licenses Gracenote Data to AI/ML developers for integration into consumer-facing products, including to ground AI outputs for consumption by end users.

99. Gracenote licenses such Gracenote Data under contractual terms that are calibrated to reduce the risk that Gracenote Data will be used improperly to compete with Gracenote and otherwise disintermediate or undermine Gracenote's core business. These restrictions include prohibitions on using Gracenote Data to recreate or replace Gracenote's products, augmenting third-party catalogs with missing fields and content from Gracenote, training foundational AI models on Gracenote content or redistributing Gracenote Data as a grounding source, and limiting the media platforms and devices Gracenote Data can be used on. These restrictions prohibit uses that would recreate the functionality of Gracenote's licensed offerings or facilitate redistribution of Gracenote Data.

100. Because Gracenote licenses its data for use in grounding AI products, any unlicensed distribution of Gracenote Data by a competing AI tool competes with Gracenote's licensed uses. Defendants' models are offered to the general public via web apps and to enterprises via API, enabling both direct end-user interactions and programmatic integration—using channels that directly compete with the consumer-facing AI grounding use that is licensed.

101. Defendants' harm to Gracenote is already manifest. Defendants' models can generate exact TMSID formats and associations, including structural recall of TMSIDs, and reproduce Gracenote's descriptive/editorial content. The models' memorization and regurgitation of Gracenote's editorial content and their structural recall of TMSID-schemas demonstrate encoding of Gracenote's expressive and organizational choices, enabling outputs that replicate

Gracenote's licensed deliverables without the paid access, quality controls, or usage constraints that preserve the market.

102. Defendants' unlicensed copying, training, and grounding of Gracenote Data erode this very market. Defendants divert demand from paid licenses by offering functionally similar outputs—identifiers, descriptive/editorial content, and cross-catalog associations—without compensation to Gracenote. And worse, Defendants negate the contractual protections that Gracenote requires of legitimate AI licensees. By operating outside these restrictions, Defendants enable the very practices that Gracenote's licenses are designed to prevent, displacing paid access and threatening downstream redistribution and re-creation at an enterprise scale.

103. Gracenote's licensing program is specifically structured to mitigate these risks through field-of-use limitations, technical controls, and downstream restrictions aligned to consumer-facing grounding. Defendants' conduct, by contrast, bypasses these safeguards and makes Gracenote-derivative outputs available through channels—consumer apps and enterprise APIs—that mirror Gracenote's licensed distribution pathways, thereby causing immediate competitive harm and long-term market erosion.

COUNT I: Copyright Infringement (17 U.S.C. § 501)

Against All Defendants

104. Gracenote incorporates by reference and realleges the preceding allegations as though fully set forth herein.

105. Gracenote is the rightful and lawful legal or beneficial owner of the copyrights in and to their works. As the owner of the registered copyright in the Gracenote Programs Database, Gracenote holds the exclusive rights to the Gracenote Programs Database and its component parts under 17 U.S.C. § 106.

106. By building training datasets containing unauthorized copies of Gracenote Data or

portions thereof, Defendants have directly infringed Gracenote's exclusive rights in its copyrighted works.

107. By storing and processing the training datasets containing unauthorized copies of Gracenote Data to train the Defendants' LLMs, including the GPT models, Defendants have directly infringed Gracenote's exclusive rights in its copyrighted works. On information and belief, by storing, training, processing, and distributing Gracenote Data, Defendants have jointly directly infringed Gracenote's exclusive rights in its copyrighted works.

108. On information and belief, Defendants' infringing conduct alleged herein was and continues to be willful and carried out with full knowledge of Gracenote's rights in the copyrighted works. As a direct result of their conduct, Defendants have wrongfully profited from copyrighted works that they do not own.

109. By and through the actions alleged above, Defendants have infringed and will continue to infringe Gracenote's exclusive rights in its copyrighted works.

110. As a direct and proximate result of Defendants' infringing conduct alleged herein, Gracenote has sustained and will continue to sustain substantial, immediate, and irreparable injury for which there is no adequate remedy at law. Unless Defendants' infringing conduct is enjoined by this Court, Defendants have demonstrated an intent to continue to infringe the copyrighted works. Gracenote therefore is entitled to permanent injunctive relief restraining and enjoining Defendants' ongoing infringing conduct.

111. Gracenote is further entitled to recover statutory damages, actual damages, restitution of profits, attorneys' fees, and other remedies provided by law.

COUNT II: Vicarious Copyright Infringement

Against OpenAI Foundation, OpenAI GP, LLC, OpenAI OpCo, LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, and OpenAI Global, LLC

112. Gracenote incorporates by reference and realleges the preceding allegations as though fully set forth herein.

113. Defendants OpenAI Foundation, OpenAI OpCo, LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC controlled, directed, and profited from the infringement perpetrated by Defendants OpenAI Global, LLC and OpenAI, LLC, including the copying and reproduction of Gracenote Data.

114. Defendants OpenAI Global, LLC and OpenAI OpCo, LLC directed, controlled, and profited from the infringement perpetrated by Defendant OpenAI, LLC, including the copying and reproduction of Gracenote Data.

115. Defendants OpenAI Foundation, OpenAI GP, LLC, OpenAI OpCo, LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, and OpenAI Global, LLC are vicariously liable for copyright infringement.

COUNT III: Contributory Copyright Infringement

Against All Defendants

116. Gracenote incorporates by reference and realleges the preceding allegations as through fully set forth herein.

117. In the alternative, to the extent an end-user may be liable as a direct infringer based on output of the GPT-based products, Defendants materially contributed to and directly assisted with the direct infringement perpetrated by end-users of the GPT-based products by way of (i) jointly-developing LLM models capable of distributing unlicensed copies of Gracenote Data to end-users; (ii) building and training the GPT LLMs using Gracenote Data; and (iii) deciding what content is actually outputted by the AI products, such as grounding output in Gracenote Data through retrieval augmented generation, fine-tuning the models for desired outcomes, and/or selecting and weighting the parameters of the GPT LLMs.

118. Defendants knew or had reason to know of the direct infringement by end-users because Defendants undertake extensive efforts in developing, testing, and troubleshooting their LLM models and GPT-based products. Defendants are fully aware that their GPT-based products are capable of distributing unlicensed copies or derivatives of copyrighted Gracenote Data.

COUNT IV: Unjust Enrichment

Against All Defendants

119. Gracenote incorporates by reference and realleges the preceding allegations as though fully set forth herein.

120. Defendants included Gracenote Data within the training corpuses for its LLMs.

121. Defendants are liable under common law principles of unjust enrichment for their reliance on Gracenote Data to train its models.

122. On information and belief, at all relevant times, Defendants have been enriched through their reliance on Gracenote Data for model training. Gracenote makes enormous investments in human talent, technology, and infrastructure to produce high-quality data. Yet without paying anything to Gracenote, Defendants exploited Gracenote Data for commercial purposes, thereby benefitting from Gracenote's extensive production efforts.

123. These OpenAI models, which were developed with Gracenote Data, now power lucrative user-facing AI products and features that Defendants continue to develop—including ChatGPT—which are critical for OpenAI's ongoing success.

124. Defendants have monetized these products with great commercial success. For example, OpenAI charges users subscription fees to access ChatGPT Plus, ChatGPT Pro, and ChatGPT Enterprise, and have earned substantial revenues as a result.

125. Defendants' enrichment has come at Gracenote's expense. Defendants' conduct diminishes the commercial value of Gracenote Data. If Defendants can exploit Gracenote Data for

commercial purposes without paying a dime to Gracenote, other companies will have fewer incentives to pay Gracenote a fair price, or any price, for Gracenote Data.

126. While models may in some instances "memorize" training works by encoding retrievable copies in their parameters, many training works are not memorized in this way. Likewise, while model outputs often may be substantially similar to works on which they are grounded, often they are not. The tuning of models that does not result in the creation of memorized copies of training works in the model parameters and the presentation of model outputs that are not substantially similar to works on which those outputs are grounded are distinct acts of exploitation that are not preempted by the Copyright Act.

127. Given these circumstances, equity and good conscience require restitution to Gracenote. Defendants should be ordered to pay Gracenote a fair price for using Gracenote Data to train and ground its models and/or disgorge to Gracenote the profits that Defendants earned from its misconduct.

128. Defendants' conduct has injured Gracenote, and Gracenote is entitled to restitution and/or disgorgement of profits and other remedies provided by law.

V. PRAYER FOR RELIEF

WHEREFORE, Gracenote demands judgment against each Defendant as follows:

1. Awarding Gracenote statutory damages, compensatory damages, restitution, disgorgement, and any other relief that may be permitted by law or equity;

2. Permanently enjoining Defendants from the unlawful, unfair, and infringing conduct alleged herein;

3. Ordering destruction under 17 U.S.C. § 503(b) of all GPT or other LLM models and training sets that incorporate Gracenote Data;

4. An award of costs, expenses, and attorneys' fees as permitted by law; and

5. Such other or further relief as the Court may deem appropriate, just, and equitable.

VI. JURY DEMAND

Gracenote hereby demands a jury trial for all claims so triable.

Dated: March 10, 2026

SUSMAN GODFREY, L.L.P.

/s/ *Elisha Barron*

Ian B. Crosby (*pro hac vice forthcoming*)
Steven M. Seigel (*pro hac vice forthcoming*)
SUSMAN GODFREY L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
Telephone: (206) 505-3843
Facsimile: (206) 516-3883
Icrosby@susmangodfrey.com
sseigel@susmangodfrey.com

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

Charisma Nguepdo (*pro hac vice forthcoming)*
Elise Miller (*pro hac vice forthcoming*)
SUSMAN GODFREY L.L.P.
1000 Louisiana St., Suite 5100
Houston, TX 77002
Telephone: (713) 651-9366
Facsimile: (713) 654-6666
cnguepdo@susmangodfrey.com
elmiller@susmangodfrey.com

Elisha Barron (EB6850)
SUSMAN GODFREY L.L.P.
One Manhattan West, 50th Fl.
New York, NY 10001
Telephone: (212) 336-8330
Facsimile: (212) 336-8340
ebarron@susmangodfrey.com

Attorneys for Plaintiff