BURSOR & FISHER, P.A.
L. Timothy Fisher (State Bar No. 191626)
Joshua B. Glatt (State Bar No. 354064)
1990 North California Blvd., 9th Floor
Walnut Creek, CA 94596
Telephone: (925) 300-4455
Facsimile: (925) 407-2700
E-mail: ltfisher@bursor.com
jglatt@bursor.com

BURSOR & FISHER, P.A.
Joseph I. Marchese (*pro hac vice* forthcoming)
Julian C. Diamond (*pro hac vice* forthcoming)
1330 Avenue of the Americas, 32nd Floor
New York, NY 10019
Telephone: (646) 837-7150
Facsimile: (212) 989-9163
E-Mail: jmarchese@bursor.com
jdiamond@bursor.com

Attorneys for Plaintiffs

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

DAVID MILLETTE and RUSLANA PETRYAZHNA, individually and on behalf of all others similarly situated,

Plaintiffs,

v.

OPENAI, INC., and OPENAI OPCO, L.L.C.,

Defendants.

Case No. 5:24-cv-04710-EJD

FIRST AMENDED CLASS ACTION COMPLAINT

JURY TRIAL DEMANDED

Plaintiffs David Millette and Ruslana Petryazhna, (collectively, "Plaintiffs"), bring this action on behalf of themselves and all others similarly situated against Defendants OpenAI, Inc. and OpenAI OpCo, L.L.C. (collectively, "OpenAI" or "Defendants"). Plaintiffs seek to recover injunctive relief and damages as a result of Defendants' unlawful conduct. Plaintiffs make the following allegations pursuant to the investigation of their counsel and are based upon information and belief, except as to the allegations specifically pertaining to themselves, which are based on personal knowledge.

NATURE OF THE CASE

1. ChatGPT is a software product created, maintained, and sold by OpenAI.

2. ChatGPT is currently powered by artificial-intelligence (hereinafter "AI") software programs called GPT-3.5, GPT-4, and GPT-4o, that are also known as "large language models" ("LLM(s)"). A large language model is "trained" by copying massive amounts of text and extracting expressive information from it. This body of text is called the "training dataset." Once a large language model has copied and ingested the text in its training dataset, it is able to emit convincingly naturalistic text outputs in response to user prompts.

3. Large language models' output is therefore entirely and uniquely reliant on the material in their training dataset. Every time they assemble text, video or image outputs, the models rely on information extracted from their training dataset.

4. This case addresses the surreptitious, non-consensual transcription of millions of YouTube users' videos by Defendants to train Defendants' AI software products. For years, YouTube has been a popular video sharing platform that allows content creators and users to upload and share videos with audiences worldwide. However, unbeknownst to those who upload videos to YouTube, Defendants have been covertly transcribing YouTube videos to create training datasets that they then use to train their AI products.

5. Plaintiffs and Class Members are YouTube users and music and video creators. Plaintiffs and Class Members have retained ownership rights in their uploaded videos, per YouTube's Terms of Service. Plaintiffs and Class Members did not consent to the use of their

videos as training material for ChatGPT. Nonetheless, their materials were transcribed and used to train ChatGPT.

6. By transcribing and using these videos in this way, Defendants profit from Plaintiffs' and Class Members' data time and time again. As Defendants' AI products become more sophisticated through the use of training datasets, they become more valuable to prospective and current users, who purchase subscriptions to access Defendants' AI products.

7. By collecting and using this data without consent, Defendants have profited significantly from the use of Plaintiffs' and Class Members' materials, violated California's Unfair Competition Law ("UCL"), infringed on Plaintiffs' ownership rights in their works, and have been unjustly enriched at Plaintiffs' and Class Members' expense.

JURISDICTION AND VENUE

8. This Court has subject matter Jurisdiction over this action pursuant to the Class Action Fairness Act ("CAFA"), 28 U.S.C. § 1332(d)(2), because this is a class action in which at least one member of the class is a citizen of a state different from any Defendants, the amount in controversy exceeds $5 million, exclusive of interest and costs, and the proposed class contains more than 100 members.

9. This Court has personal jurisdiction over the Defendants because Defendants maintain their principal places of business in this District and because a substantial part of the events or omissions giving rise to the claims asserted herein occurred in this District.

10. Venue is proper in this district pursuant to 28 U.S.C. § 1391 because a substantial part of the events or omissions giving rise to the claims asserted herein occurred in this District and because Defendants maintain their principal places of business in this District.

PARTIES

11. Plaintiff David Millette is a citizen of Massachusetts and resident of Douglas, Massachusetts. Plaintiff created a YouTube account in or around June 2009. During that entire time, Plaintiff Millette has retained ownership rights to the video content he has uploaded to YouTube, per YouTube's Terms of Service.

12. Plaintiff Ruslana Petryazhna is a citizen of New York and resident of Astoria, New York. Plaintiff Petryazhna created a YouTube account in or around 2008. During that entire time, Plaintiff Petryazhna has retained ownership rights to the video content and numerous songs she has uploaded to YouTube, per YouTube's Terms of Service. Moreover, Plaintiff Petryazhna has registered her copyright of those works with the United States Copyright Office, *see Ex. A*, including her song "A Bubble World," which was uploaded to YouTube 15 years ago. *See Ex. B*.

13. Defendants transcribed Plaintiffs' videos to train their AI software products.

14. Defendant OpenAI, Inc. is a Delaware nonprofit corporation with its principal place of business located at 3180 18th Street, San Francisco, CA 94110.

15. Defendant OpenAI OpCo, L.L.C. is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI OpCo, L.L.C. is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI, Inc. controls OpenAI OpCo, L.L.C. directly and through other OpenAI entities.

16. Each Defendant acted jointly to perpetrate the acts described herein. At all times relevant to the allegations in this matter, each Defendant acted in concert with, with the knowledge and approval of, and/or as the agent of, the other Defendant within the course and scope of the agency, regarding the acts and omissions alleged.

GENERAL BACKGROUND

17. OpenAI creates and sells artificial-intelligence ("AI") software products. AI software is designed to algorithmically simulate human reasoning or inference, often using statistical methods.

18. Certain AI products created and sold by OpenAI are known as "large language models." Large language models ("LLMs") are types of AI software designed to parse and emit natural language. Though LLMs are software programs, they are not created the way most software programs are—that is, by human software engineers writing code. Rather, LLMs are "trained" by copying massive amounts of text from various sources and feeding these copies into the model. During training, these models copy each piece of information in the training dataset and extract expressive information from it. The LLMs progressively adjust their output to more closely

resemble the images, videos, and sequences of words copied from the training dataset. Once these models have copied and ingested all these inputs, they are able to emit convincing simulations of natural written language, as well as videos and images as they appear in the training dataset.

19. As the U.S. Patent and Trademark Office has observed, LLM "training" "almost by definition involve[s] the reproduction of entire works or substantial portions thereof."[1]

20. Much of the material in OpenAI's training datasets, however, comes from works—including videos created and uploaded by Plaintiffs—that were copied by OpenAI without consent, without credit, and without compensation.

21. Specifically, OpenAI made a series of LLMs, including without limitation GPT-1 (released June 2018), GPT-2 (February 2019), GPT-3 (May 2020), GPT-3.5 (March 2022), GPT-4 (March 2023) and GPT-4o (May 2024). "GPT" is an abbreviation for "generative pre-trained transformer," where "pre-trained" refers to the use of textual material for training, "generative" refers to the model's ability to emit text, and "transformer refers" to the underlying training algorithm. OpenAI offers certain language models in variant forms. For instance, the GPT-4 family of models includes publicly accessible variants called "gpt-4-0125-preview," "gpt-4-turbo-preview," and "gpt-4-32k;" and the GPT-3.5 Turbo family of models includes publicly accessible variants called "gpt-3.5-turbo-0125," "gpt-3.5-turbo-1106," and "gpt-3.5-turbo-instruct." OpenAI has made other language-model variants that are in commercial use but are not publicly accessible. In an interview with the Financial Times in November 2023, OpenAI CEO Sam Altman confirmed that GPT-5 is under development. Together, OpenAI's LLMs, including any in development, will be referred to as the "OpenAI Language Models."[2]

22. Many kinds of material have been used to train large language models. Video transcriptions, however, are a key ingredient in training datasets for LLMs because they offer copious examples of natural language.

---

[1] U.S. Patent & Trademark Office, *Public Views on Artificial Intelligence and Intellectual Property Policy*, 2020, *available* https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf.

[2] The definition of "OpenAI Language Models" encompasses any language models developed (or in development) by OpenAI, irrespective of whether those models underly ChatGPT.

23. In 2022, OpenAI released an automatic speech recognition ("ASR") system called "Whisper." The Whisper model, which transcribes audio into text, was trained on 680,000 hours of data collected from across the web. Tellingly, the exact names of the speech recognition corpora on which Whisper was trained are unavailable. But one of the world's largest open multilingual speech corpora, VoxPopuli, contains only 400,000 hours of unlabeled speech data. Libriheavy, an ASR corpus considered one of the largest freely available corpora of speech with supervisions, only consists of 50,000 hours of English speech derived from LibriVox.

24. There are only a handful of publicly available, internet-based speech corpora that can be utilized as training data for LLMs. As demonstrated, the two biggest corpora combined (VoxPopuli and Libriheavy) still fall more than 200,000 hours short of the duration of speech that comprises Whisper's training dataset.

25. The New York Times reported that Whisper is capable of transcribing the audio from YouTube videos, and that an OpenAI team that included OpenAI's president, Greg Brockman, transcribed more than one million hours of video from YouTube.[3] This tracks with OpenAI's admission to the U.S. Patent and Trademark Office that the company produces AI products that "draw on … experience developing cutting-edge technical AI systems, including by the use of large, publicly available datasets that include copyrighted works."[4]

26. Defendants could have "trained" their LLMs using works in the public domain. They could have paid a reasonable licensing fee to use those works. What Defendants could not do is ignore protections afforded by virtue of Plaintiffs' ownership rights in their respective works. As evident by congressional testimony, Defendants' CEO, Sam Altman, acknowledged that "[a]s for

[3] Cade Metz, et al., *How Tech Giants Cut Corners to Harvest Data for A.I.*, THE NEW YORK TIMES (Apr. 6, 2024) https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html.

[4] OpenAI, *Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, U.S. Pat. & Trademark Off., Dkt. No. PTO-C-2019-0038, *available* chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf.

enabling artist control, OpenAI's training data comes from a variety of sources, including … also by licensing content directly from content owners."[5] Just not Plaintiffs here.

27. OpenAI's Language Models' datasets include transcriptions of videos taken directly from YouTube, because these video transcriptions are one of the largest corpora of natural language data available for training and fine-tuning the OpenAI Language Models.

28. Moreover, by directing its Langue Models to scrape videos hosted on YouTube—a site that expressly grants the uploader retention in their ownership rights of the videos—Defendants know, or should have known, of a strong likelihood that copyrighted content would be included in its scraping.

29. In fact, ChatGPT offers an analysis of the works. ChatGPT's analysis of "A Bubble World" explains that "[t]he tone is light-hearted, reminiscent of childhood innocence, emphasizing the importance of play and joy in life" and that the "lyrics typically reflect a longing of freedom and joy, contrasting with the constraints of everyday life." This suggests that the underlying LLM must have ingested the entire song during its "training."

30. Notably, Mr. Altman knows that OpenAI's unauthorized use of a creator's content for the company's benefit is problematic. In testimony to Congress, Mr. Altman explained that "[e]nsuring that the creator economy continues to be vibrant is an important priority for OpenAI. … OpenAI does not want to replace creators. We want our systems to be used to empower creativity, and to support and augment the essential humanity of artists and creators."[6] And, in a subsequent interview, Mr. Altman explained that "creators deserve control over how their creations are used."[7]

[5] Sam Altman, *Questions for the Record*, at 10 (June 22, 2023), *available* https://www.judiciary.senate.gov/imo/media/doc/2023-05-16_-_qfr_responses_-_altman.pdf.

[6] *Id*.

[7] Ted Johnson, *OpenAI CEO Sam Altman Says Content Owners Need To Get 'Significant Upside Benefit' From New Technology*, DEADLINE (May 16, 2023) *available* https://deadline.com/2023/05/ai-chat-gpt-senate-sam-altman-1235368420/.

CLASS ALLEGATIONS

31. Plaintiffs seek to represent a class defined as all persons or entities domiciled in the United States who uploaded any YouTube video that was transcribed and then used as training data for the OpenAI Language Models without their consent (the "Nationwide Creator Class").

32. Plaintiff Millette seeks to represent a class defined as all persons or entities domiciled in Massachusetts that uploaded any YouTube video that was transcribed and then used as training data for the OpenAI Language Models without their consent (the "Massachusetts Creator Subclass") (collectively with the Nationwide Creator Class, the "Creator Classes").

33. Plaintiff Petryazhna seeks to represent a class defined as all persons or entities domiciled in the United States whose registered copyright material within any YouTube video was transcribed and then used as training data for the OpenAI Language Models without their consent (the "Copyright Class").

34. The Creator Classes and Copyright Class together shall be referred to as the "Classes."

35. Specifically excluded from the Classes are Defendants, Defendants' officers, directors, agents, trustees, parents, children, corporations, trusts, representatives, employees, principals, servants, partners, joint ventures, or entities controlled by Defendants, and their heirs, successors, assigns, or other persons or entities related to or affiliated with Defendants and/or Defendants officers and/or directors, the judge assigned to this action, and any member of the judge's immediate family.

36. Plaintiffs reserve the right to expand, limit, modify, or amend the class definitions, including the addition of one or more subclasses, in connection with their motion for class certification, or at any other time, based on, *inter alia*, changing circumstances and/or new facts obtained.

37. **Numerosity**. On information and belief, thousands of individuals fall into the definitions of the Classes. Members of the Classes can be identified through Defendants' records, discovery, and other third-party sources.

38. **Commonality and Predominance**. Common questions of law and fact exist as to all Members of the Classes and predominate over any questions affecting only individual Members of the Classes. These common legal and factual questions include, but are not limited to, the following:

a. Whether Defendants violated the rights of Plaintiffs and the Members of the Classes when they transcribed Plaintiffs' videos and used those transcriptions as part of their AI software's training datasets;

b. Whether Defendants' conduct violated the rights and protections afforded to holders of registered copyrights;

c. Whether Defendants' conduct alleged herein constitutes Unfair Competition under California Business and Professions Code §§ 17200 *et seq.*;

d. Whether this Court should enjoin Defendants from engaging in the unlawful conduct alleged herein, and what the scope of that injunction would be;

e. Whether any affirmative defense excuses Defendants' conduct;

f. Whether any statutes of limitation constrain the potential recovery for Plaintiffs and the Classes;

g. Whether Plaintiffs and the other Creator and Copyright Class Members are entitled to restitution or other relief; and

h. Whether Defendants' conduct alleged herein constitutes a violation of Mass. Gen. Law Ch. 93A.

39. **Typicality**. Plaintiffs' claims are typical of the claims of the other Members of the respective Classes they seek to represent in that, among other things, all Class Members were similarly situated and were comparably injured through Defendants' wrongful conduct as set forth herein. Further, there are no defenses available to Defendants that are unique to Plaintiffs and not applicable to the respective Classes they seek to represent.

40. **Adequacy of Representation**. Plaintiffs will fairly and adequately protect the interests of the Classes. Plaintiffs have retained counsel that is highly experienced in complex

class action litigation, and Plaintiffs intends to vigorously prosecute this action on behalf of the Classes. Furthermore, Plaintiffs have no interests that are antagonistic to those of the Classes.

41. **Superiority**. A class action is superior to all other available means for the fair and efficient adjudication of this controversy. The damages or other financial detriment suffered by individual Class Members are relatively small compared to the burden and expense of individual litigation of their claims against Defendants. It would thus be virtually impossible for the Classes to obtain effective redress for the wrongs committed against the Members on an individual basis. Furthermore, even if Class Members could afford such individualized litigation, the court system could not. Individualized litigation would create the danger of inconsistent or contradictory judgments arising from the same set of facts. Individualized litigation would also increase the delay and expense to all parties and the court system from the issues raised by this action. By contrast, the class action device provides the benefits of adjudication of these issues in a single proceeding, economies of scale, and comprehensive supervision by a single court, and presents no unusual management difficulties under the circumstances.

42. Further, Defendants have acted and refused to act on grounds generally applicable to the proposed Classes, thereby making appropriate final injunctive and declaratory relief with respect to the Classes as a whole.

CAUSES OF ACTION

COUNT I
Unjust Enrichment or Restitution
(On behalf of Plaintiffs and the Classes)

43. Plaintiffs incorporate by reference and re-allege each and every allegation set forth above as though fully set forth herein.

44. Plaintiffs bring this claim under the laws of the State of California.

45. Plaintiffs bring this claim individually and on behalf of Members of the Classes against the Defendants.

46. To the extent required by law, Plaintiffs bring this claim in the alternative to any legal claims that may be alleged.

47. Plaintiffs also alternatively allege this claim as a Quasi-Contract or Non-Quasi-Contract Claim for Restitution and Disgorgement.

48. Plaintiffs and Members of the Classes unwittingly conferred a benefit upon Defendants. OpenAI misappropriated valuable information from Plaintiffs' and Class Members' videos to expand their AI software's training datasets and used that information to develop and improve their products. In using Plaintiffs' information to refine their Language Models, OpenAI made their products more valuable to prospective and current users, who purchase subscriptions to access them. Plaintiffs and Class Members received nothing from this transaction. Plaintiffs lack an adequate remedy at law, and plead this cause of action in the alternative to the extent Plaintiffs are required to do so.

49. Defendants have knowledge of such benefits.

50. Defendants have been unjustly enriched in retaining the revenues derived from the sales of their products trained on Plaintiffs' and Class Members' videos. Retention of those moneys under these circumstances is unjust and inequitable because Defendants did not obtain the meaningful consent of Plaintiffs and Class Members before using their videos as described above.

51. Because Defendants' retention of the non-gratuitous benefits conferred on it by Plaintiffs and Class Members is unjust and inequitable, Defendants must pay restitution to Plaintiffs and the Class Members for their unjust enrichment, as ordered by the Court.

52. Plaintiffs and the Members of the Classes lack an adequate remedy at law to address the unfair conduct at issue here. Legal remedies available to Plaintiffs and Class Members are inadequate because they are not equally prompt and certain and in other ways efficient as equitable relief. Damages are not equally certain as restitution because the standard that governs restitution is different than the standard that governs damages. Hence, the Court may award restitution even if it determines that Plaintiffs fail to sufficiently adduce evidence to support an award of damages. Damages and restitution are not the same amount. Unlike damages, restitution is not limited to the amount of money a defendant wrongfully acquired plus the legal rate of interest. Equitable relief, including restitution, entitles the plaintiff to recover all profits from the wrongdoing, even where the original funds taken have grown far greater than the legal rate of interest would recognize.

Legal claims for damages are not equally certain as restitution because claims for restitution entail few elements. In short, significant differences in proof and certainty establish that any potential legal claim cannot serve as an adequate remedy at law. Plaintiffs and Members of the putative Classes seek non-restitutionary disgorgement of the financial profits that Defendants obtained as a result of their unjust conduct.

COUNT II
UCL – Unfair Competition
Cal. Bus. & Prof. Code §§ 17200 *et seq.*
(On behalf of Plaintiffs and the Classes)

53. Plaintiffs incorporate by reference and re-allege each and every allegation set forth above as though fully set forth herein.

54. Defendants engaged in unfair business practices by, among other things, using Plaintiffs' videos to train their Language Models without permission from Plaintiffs or Class Members.

55. The unfair business practices described herein violate California Business and Professions Code §§ 17200 *et seq.* (the "UCL").

56. The unfair business practices described herein violate the UCL because they are unfair, immoral, unethical, oppressive, unscrupulous, or injurious to consumers, and because Defendants used Plaintiffs' videos for Defendants' own commercial profit without the authorization of Plaintiffs or the Classes. Defendants unfairly profit from and take credit for developing a commercial product based on unattributed reproductions of those stolen videos and content.

57. Each Plaintiff suffered economic injury as a result of Defendants' actions. Each Plaintiff was deprived of the value of their works by being deprived of the right to control who can use their works for commercial gain. By doing so, Defendants deprived Plaintiffs moneys that would be owed to them had they consented—or granted a license too—use the works.

58. Conduct is **unfair** under the UCL if it is immoral, unethical, oppressive, or unscrupulous and the conduct outweighs any benefits to consumers. Defendants' conduct was unfair because it relied on non-consensual use of Plaintiffs' works for Defendants' gain.

Defendants could have sought consent to do so but chose not to. The conduct outweighs any benefits to consumers. Defendants knowingly and secretively trained ChatGPT using unauthorized transcriptions and copies of Plaintiffs' videos. Defendants deceptively marketed their product in a manner that fails to attribute the success of their product to the work on which it is based.

COUNT III
Massachusetts Unfair and Deceptive Business Practices Act
Mass. Gen. Law Ch. 93A *et seq.*
(On Behalf of Plaintiff Millette and the Massachusetts Creator Subclass)

59. Plaintiff Millette incorporates by reference and re-alleges each and every allegation set forth above as if fully set forth herein.

60. Plaintiff Millette brings this claim individually and on behalf of the Members of the Massachusetts Subclass against Defendants.

61. Section 2 of Chapter 93—the Massachusetts Unfair and Deceptive Business Practices Act ("93A")—prevents the use of "unfair or deceptive acts or practices in the conduct of any trade or commerce."

62. Section 9 of Chapter 93 provides: "Any person … who has been injured by another person's use or employment of any method, act or practice declared to be unlawful by section two … may bring an action in the superior court … for damages and such equitable relief, including an injunction, as the court deems to be necessary and proper … Any persons entitled to bring such action may, if the use or employment of the unfair or deceptive act or practice has caused similar injury to numerous other persons similarly situated and if the court finds in a preliminary hearing that he adequately and fairly represents such other persons, bring the action on behalf of himself and such other similarly injured and situated persons."

63. Pursuant to the definitions codified in Chapter 93A § 1, each Defendant is a "person," and Defendants engaged in "trade" and "commerce" in Massachusetts by engaging in the sale of Products that directly or indirectly affect the people of Massachusetts by scraping YouTube videos created by Plaintiff and the Subclass to train ChatGPT for use by Massachusetts residents.

64. By engaging in the acts and omissions alleged above and incorporated herein, Defendants have engaged and continue to engage in unfair and deceptive acts or practices in the conduct of trade or commerce.

65. Defendants' acts deceive, or have a tendency to deceive, a reasonable consumer of the general public.

66. Defendants' acts and omissions are material, in that a reasonable person would attach importance to the information described above and would be induced to act on the information in deciding to use services such as Defendants' GPT service.

67. Plaintiff Millette and Members of the Massachusetts Subclass were deceived by Defendants' actions in that they had no idea their YouTube content was being used to train Defendants' products for Defendants' commercial benefit.

68. Plaintiff Millette and Members of the Massachusetts Subclass did not consent to Defendants' use of their content.

69. Defendants knowingly committed the acts alleged herein.

70. Had Plaintiff Millette and the Massachusetts Creator Subclass Members known that the Defendants were scraping their content for use to train their products, they would have requested compensation for the misappropriation of their works.

71. Plaintiff Millette and the Massachusetts Subclass Members were injured as a direct and proximate result of Defendants misappropriated and non-consensual use of their works.

72. Plaintiff Millette and the Massachusetts Subclass Members have been harmed by this injury, adverse consequences, and/or loss.

73. 93A represents a fundamental public policy of the Commonwealth of Massachusetts.

74. For each loss, Plaintiff Millette and each Member of the Massachusetts Subclass may recover an award of actual damages or twenty-five dollars, whichever is greater. Ch. 93A § 9(3).

75. Disgorgement of profit derived from an unfair and deceptive act or practice is a permissible damage remedy under G.L. c. 93A, § 9.

76. Accordingly, Plaintiff Millette and the Members of the Massachusetts Subclass seek the disgorgement of profits that Defendants have derived from the use of their works to train Defendants' product.

77. Because Defendants acted willfully or knowingly, Plaintiff Millette and the Massachusetts Subclass Members may recover up to three, but not less than two times, this amount. In addition, Plaintiff Millette may recover attorneys' fees and costs.

78. Plaintiff Millette and each Member of the Massachusetts Subclass may recover an award of actual damages (in this case unlawful profit by Defendants made possible by training their LLMs with Plaintiff's and Class Members' works) or twenty-five dollars, whichever is greater. Ch. 93A § 9(3).

79. Plaintiff Millette and the Members of the Massachusetts Subclass may also seek the imposition of injunctive relief which limits and polices Defendants' actions reaching Massachusetts. The balance of the equities favors the entry of permanent injunctive relief against Defendants. Plaintiff, Members of the Massachusetts Subclass, and the general public will be irreparably harmed, absent the entry of permanent injunctive relief against Defendants. Plaintiff Millette and Members of the Massachusetts Subclass and the general public lack an adequate remedy at law. A permanent injunction against Defendants is in the public interest. Defendants' unlawful behavior is capable of repetition or re-occurrence absent the entry of a permanent injunction.

COUNT IV
Direct Copyright Infringement
17 U.S.C. § 501
(On Behalf of Plaintiff Petryazhna and the Copyright Class)

80. Plaintiff Petryazhna incorporates by reference and re-alleges each and every allegation set forth above as if fully set forth herein.

81. As the rightful and legal owner of the registered copyrights in videos that were transcribed and then used to train the OpenAI Language Models, Plaintiff Petryazhna holds copyright rights to those videos under 17 U.S.C. § 106.

82. Plaintiff Petryazhna and the Copyright Class Members' works are original to their creators and are fixed in tangible mediums of expression under 17 U.S.C. § 102(a)(2), (6), and (7).

83. Plaintiff Petryazhna and the Members of the Copyright Class have duly and timely registered their copyrights in their works with the U.S. Copyright Office.

84. Plaintiff Petryazhna and the Copyright Class Members are legal or beneficial owners of the exclusive right to reproduce their copyrighted works in copies under 17 U.S.C. § 106(1) and (2), as well as the right to refrain from such reproduction.

85. Plaintiff never authorized OpenAI to transcribe and make copies of her videos, make derivative works, publicly display copies (or derivative works), or distribute copies (or derivative works). All those rights belong exclusively to Plaintiff Petryazhna and the Members of the Copyright Classes under copyright law.

86. To train the OpenAI Language Models, OpenAI relied on transcribing and harvesting mass quantities of YouTube videos from the public internet, including Plaintiff Petryazhna's videos.

87. OpenAI transcribed and made copies of Plaintiff's videos during the training process of the OpenAI Language Models without Plaintiff's permission. Because the OpenAI Language Models cannot function without the expressive information extracted from Plaintiff Petryazhna's works (and others) and retained inside them, the OpenAI Language Models are themselves infringing derivative works, made without Plaintiff Petryazhna's permission and in violation of her exclusive rights under the Copyright Act.

88. Plaintiff Petryazhna has been injured by OpenAI's acts of direct copyright infringement. Plaintiff Petryazhna is entitled to statutory damages, actual damages, restitution of profits, and/or other remedies provided by law.

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs, individually and on behalf of all others similarly situated, seek judgment against Defendants, as follows:

a. For an order certifying the Classes under Rule 23 of the Federal Rules of Civil Procedure and naming Plaintiffs as the representative for the Classes and Plaintiffs' attorneys as Class Counsel;

b. For an order declaring that Defendants' conduct violates the laws referenced herein;

c. For an order finding in favor of Plaintiffs and the Classes on all counts asserted herein;

d. For compensatory, statutory, and punitive damages in amounts to be determined by the Court and/or jury;

e. For prejudgment interest on all amounts awarded;

f. For an order of restitution and all other forms of equitable monetary relief;

g. For injunctive relief as the Court may deem proper; and

h. For an order awarding Plaintiffs and the Classes their reasonable attorneys' fees and expenses and costs of suit.

DEMAND FOR TRIAL BY JURY

Pursuant to Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury of any and all issues in this action so triable of right.

Dated: October 18, 2024

BURSOR & FISHER, P.A.

By: /s/ *L. Timothy Fisher*
L. Timothy Fisher

L Timothy Fisher (State Bar No. 191626)
Joshua B. Glatt (State Bar No. 354064)
1990 North California Blvd., 9th Floor
Walnut Creek, CA 94596
Telephone: (925) 300-4455
Facsimile: (925) 407-2700
E-mail: ltfisher@bursor.com
jglatt@bursor.com

BURSOR & FISHER, P.A.
Joseph I. Marchese (*pro hac vice* forthcoming)
Julian C. Diamond (*pro hac vice* forthcoming)
1330 Avenue of the Americas, 32nd Floor
New York, NY 10019

Telephone: (646) 837-7150
Facsimile: (212) 989-9163
E-Mail: jmarchese@bursor.com
jdiamond@bursor.com

Attorneys for Plaintiffs