	Case 5:24-cv-04708 Document 1 File	ed 08/02/24 Page 1 of 11
1 2 3 4 5 6 7 8 9 10	 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 (<i>pro hac vice</i> forthcoming) Julian C. Diamond (<i>pro hac vice</i> forthcoming) 1330 Avenue of the Americas, 32nd Floor New York, NY 10019 Telephone: (646) 837-7150 Facsimile: (212) 989-9163 	
10	E-Mail: jmarchese@bursor.com jdiamond@bursor.com	
12	Attorneys for Plaintiff	
13		
14	UNITED STATES D	ISTRICT COURT
15	NORTHERN DISTRIC	CT OF CALIFORNIA
	DAVID MILLETTE, individually and on behalf	Case No.
16	•	Cuse Ivo.
16 17	of all others similarly situated,	CLASS ACTION COMPLAINT
	•	
17	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and	CLASS ACTION COMPLAINT
17 18	of all others similarly situated, Plaintiff, v.	CLASS ACTION COMPLAINT
17 18 19	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and	CLASS ACTION COMPLAINT
17 18 19 20	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and ALPHABET INC.,	CLASS ACTION COMPLAINT
17 18 19 20 21	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and ALPHABET INC.,	CLASS ACTION COMPLAINT
 17 18 19 20 21 22 	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and ALPHABET INC.,	CLASS ACTION COMPLAINT
 17 18 19 20 21 22 23 	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and ALPHABET INC.,	CLASS ACTION COMPLAINT
 17 18 19 20 21 22 23 24 	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and ALPHABET INC.,	CLASS ACTION COMPLAINT
 17 18 19 20 21 22 23 24 25 	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and ALPHABET INC.,	CLASS ACTION COMPLAINT
 17 18 19 20 21 22 23 24 25 26 	of all others similarly situated, Plaintiff, v. GOOGLE LLC, YOUTUBE INC., and ALPHABET INC.,	CLASS ACTION COMPLAINT

Plaintiff David Millette, (hereinafter "Plaintiff"), brings this action on behalf of himself and all others similarly situated against Defendants Google LLC, YouTube Inc., and Alphabet Inc. (collectively, "Google" or "Defendants"). Plaintiff seeks to recover injunctive relief and damages as a result of Defendants unlawful conduct. Plaintiff makes the following allegations pursuant to the investigation of his counsel and are based upon information and belief, except as to the allegations specifically pertaining to himself, which are based on personal knowledge.

NATURE OF THE CASE

1.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

Gemini is a software product created, maintained, and sold by Google.

2. Gemini currently includes five AI software programs called Gemini 1.0 Ultra, Gemini 1.5 Pro, Gemini 1.0 Pro, Gemini 1.0 Nano and Gemini 1.5 Flash, also known as *multimodal large language models* ("MLLM"). A multimodal large language model, in comparison to a large language model, is capable of processing information across modalities, meaning information can be input in the form of videos, photos, and text. Similar to a large language model, an MLLM is "trained" by copying massive amounts of text alongside images and videos, and extracting expressive information from it. This body of information is called the *training dataset*. Once an MLLM has copied and ingested the information in its training dataset, it is able to emit convincingly naturalistic text, video or photo outputs in response to user prompts.

3. Large language models' ("LLMs") and MLLMs' output are 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 the information they extracted from their training dataset.

4. This case addresses the surreptitious, non-consensual transcription and use of millions of YouTube users' videos by Defendants to train Defendants' AI software products.

5. 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 Defendants then use to train their AI products.

6. Plaintiff and Class members are YouTube users and video creators. Plaintiff and Class members have retained ownership rights in their uploaded videos, per YouTube's Terms of

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

Service. Plaintiff and Class members did not consent to the use of their videos as training material for Gemini. Nonetheless, their videos were transcribed and used to train Gemini.

7. By transcribing and using these videos in this way, Defendants profit from
Plaintiff's 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.

8. By collecting and using this data without consent, Defendants have profited significantly from the use of Plaintiff's and Class members' videos, violated California's Unfair Competition Law ("UCL"), and have been unjustly enriched at Plaintiff and Class members' expense.

JURISDICTION AND VENUE

9. 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.

10. 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.

11. 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

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

27 13. Plaintiff's videos have been transcribed by Defendants to train their AI software
28 products.

14. Defendant Google, LLC is a Delaware limited liability company with its headquarters at 1600 Amphitheatre Parkway, Mountain View, California 94093.

15. Defendant YouTube Inc. is a Delaware limited liability company with its headquarters at 901 Cherry Ave. San Bruno, CA 94066.

16. Defendant Alphabet Inc. is a Delaware limited liability company with its headquarters at 1600 Amphitheatre Parkway, Mountain View, California 94093.

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

GENERAL BACKGROUND

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

19. Certain AI products created and sold by Google are known as *multimodal large* language models. Both multimodal large language models ("MLLMs") and large language models ("LLMs") are types of AI software designed to parse and emit natural language. Though MLLMs and LLMs are software programs, they are not created the way most software programs are—that is, by human software engineers writing code. Rather, MLLMs are "trained" by ingesting large amounts of text, images and videos into the model. These corpora of input materials are called the training datasets. During training, these models copy each piece of information in the training dataset and extract expressive information from it. The MLLMs and LLMs progressively adjust their output to more closely resemble the images, videos, and sequences of words copied from the 24 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.

27

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

25

26

28

Case 5:24-cv-04708 Document 1 Filed 08/02/24 Page 5 of 11

20. Much of the material in Google's training datasets, however, comes from videos including videos created and uploaded by Plaintiff and class members-that were copied and transcribed by Google without consent, without credit, and without compensation.

21. Google made a series of multimodal large language models, including without limitation Gemini 1.0 Ultra, Gemini 1.5 Pro, Gemini 1.0 Pro, Gemini 1.0 Nano and Gemini 1.5 Flash. Google is working on other language-model variants such as Veo, a generative video model, that is not yet publicly available. Together, Google's large language models, including any in development, will be referred to as the "Google Language Models."¹

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

23. Google claims that the training datasets for their Gemini models comprise web documents, books, code, images, audio, and video data. Tellingly, there is little specificity about the source of this training data.

In 2023, Google changed its terms of service. An April 2024 New York Times 24. article reported that part of the rationale for this change was a need for more data that could be used in Gemini's training datasets.

25. Google did not properly obtain consent for its use and profiting from materials that were uploaded before this change to Google's terms of service.

26. The same New York Times report claims that at least five people with knowledge of Google's practices knew that Google transcribed YouTube videos to harvest text for its Language Models.

23 27. Google'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 Google's Language Models.

25 26

27

28

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

24

¹ The definition of "Google Language Models" encompasses any language models developed (or in development) by Google, irrespective of whether those models underly Gemini.

CLASS ACTION COMPLAINT – JURY TRIAL DEMANDED

Case 5:24-cv-04708 Document 1 Filed 08/02/24 Page 6 of 11

CLASS ALLEGATIONS

28. Plaintiff seeks to represent a class defined as all persons or entities domiciled in the United States that uploaded any YouTube video that was transcribed and then used as training data for the Google Language Models without their consent (the "Nationwide Class").

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

30. Specifically excluded from the Class 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.

31. Plaintiff reserves the right to expand, limit, modify, or amend the class definition, including the addition of one or more subclasses, in connection with his motion for class certification, or at any other time, based on, inter alia, changing circumstances and/or new facts obtained.

32. **Numerosity**. On information and belief, hundreds of thousands of video creators fall into the definitions of the Class. Members of the Class can be identified through Defendants' records, discovery, and other third-party sources.

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

 a. Whether Defendants violated the copyrights of Plaintiff and the Class when they transcribed Plaintiff's videos and used those transcriptions as part of their AI software's training datasets;

CLASS ACTION COMPLAINT – JURY TRIAL DEMANDED

	Case 5:24-cv-04708 Document 1 Filed 08/02/24 Page 7 of 11			
1 2	b. Whether Gemini is itself an infringing derivative work based on Plaintiff's videos;			
3	c. Whether Defendants' use of Plaintiff's videos to train their Language			
4	Models constitutes unjust enrichment;			
5	d. Whether Defendants' conduct alleged herein constitutes Unfair Competition			
6	under California Business and Professions Code § 17200 et seq.			
7	e. Whether this Court should enjoin Defendants from engaging in the unlawful			
8	conduct alleged herein, and what the scope of that injunction would be.			
9	f. Whether any affirmative defense excuses Defendants' conduct.			
10	g. Whether any statutes of limitation constrain the potential recovery for			
11	Plaintiff and the Class.			
12	h. Whether Plaintiff and the other Class members are entitled to restitution or			
13	other relief.			
14	34. Typicality . Plaintiff's claims are typical of the claims of the other members of the			
15	Class in that, among other things, all Class members were similarly situated and were comparably			
16	injured through Defendants' wrongful conduct as set forth herein. Further, there are no defenses			
17	available to Defendants that are unique to Plaintiff.			
18	35. Adequacy of Representation. Plaintiff will fairly and adequately protect the			
19	interests of the Class. Plaintiff has retained counsel that is highly experienced in complex			
20	consumer class action litigation, and Plaintiff intends to vigorously prosecute this action on behalf			
21	of the Class. Furthermore, Plaintiff has no interests that are antagonistic to those of the Class.			
22	36. Superiority . A class action is superior to all other available means for the fair and			
23	efficient adjudication of this controversy. The damages or other financial detriment suffered by			
24	individual Class members are relatively small compared to the burden and expense of individual			
25	litigation of their claims against Defendants. It would thus be virtually impossible for the Class to			
26	obtain effective redress for the wrongs committed against the members on an individual basis.			
27	Furthermore, even if Class members could afford such individualized litigation, the court system			
28	could not. Individualized litigation would create the danger of inconsistent or contradictory			
	CLASS ACTION COMPLAINT – JURY TRIAL DEMANDED 6			

I

Case 5:24-cv-04708 Document 1 Filed 08/02/24 Page 8 of 11

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.
37. Further, Defendants have acted and refused to act on grounds generally applicable
to the proposed Class, thereby making appropriate final injunctive and declaratory relief with

respect to the Class as a whole.

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

CAUSES OF ACTION

<u>COUNT I</u> Unjust Enrichment or Restitution (On behalf of Plaintiff and the Class)

38. Plaintiff incorporates by reference and re-alleges each and every allegation set forth above as though fully set forth herein.

39. Plaintiff brings this claim individually and on behalf of members of the Class against the Defendants.

40. To the extent required by law, Plaintiff brings this claim in the alternative to any legal claims that may be alleged.

41. Plaintiff also alternatively alleges this claim as a Quasi-Contract or Non-Quasi-Contract Claim for Restitution and Disgorgement.

42. Plaintiff and Class members unwittingly conferred a benefit upon Defendants. Google acquired valuable information from Plaintiff and Class members' videos to expand their AI software's training datasets and used that information to develop and improve their products. In using Plaintiff's information to refine its Language Models, Defendants made their products more valuable to prospective and current users, who purchase subscriptions to access them. Plaintiff and Class members received nothing from this transaction. Plaintiff lacks an adequate remedy at law, and pleads this cause of action in the alternative to the extent Plaintiff is required to do so.

- 27 28
- 43. Defendants have knowledge of such benefits.

CLASS ACTION COMPLAINT – JURY TRIAL DEMANDED

Case 5:24-cv-04708 Document 1 Filed 08/02/24 Page 9 of 11

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

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

8 46. Plaintiff and the members of the Class lack an adequate remedy at law to address 9 the unfair conduct at issue here. Legal remedies available to Plaintiff and class members are 10 inadequate because they are not equally prompt and certain and in other ways efficient as equitable 11 relief. Damages are not equally certain as restitution because the standard that governs restitution 12 is different than the standard that governs damages. Hence, the Court may award restitution even if 13 it determines that Plaintiff fails to sufficiently adduce evidence to support an award of damages. 14 Damages and restitution are not the same amount. Unlike damages, restitution is not limited to the 15 amount of money a defendant wrongfully acquired plus the legal rate of interest. Equitable relief, 16 including restitution, entitles the plaintiff to recover all profits from the wrongdoing, even where 17 the original funds taken have grown far greater than the legal rate of interest would recognize. 18 Legal claims for damages are not equally certain as restitution because claims for restitution entail 19 few elements. In short, significant differences in proof and certainty establish that any potential 20 legal claim cannot serve as an adequate remedy at law.

47. Plaintiff and members of the putative class seek non-restitutionary disgorgement of the financial profits that Defendants obtained as a result of their unjust conduct.

<u>COUNT II</u> UCL – Unfair Competition Cal. Bus. & Prof. Code §§ 17200 *et seq.* (On behalf of Plaintiff and the Class)

48. Plaintiff incorporates by reference and re-alleges each and every allegation set forth above as though fully set forth herein.

28

27

21

22

23

24

25

26

1

2

3

4

5

6

7

Case 5:24-cv-04708 Document 1 Filed 08/02/24 Page 10 of 11

49. Defendants engaged in unfair business practices by, among other things, usingPlaintiff's videos to train their Language Models without permission from Plaintiff or Classmembers.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

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

51. 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 Plaintiff's videos to train their Language Models for Defendants' own commercial profit without the authorization of Plaintiff or the Class. Defendants unfairly profit from, and take credit for, developing a commercial product based on unattributed reproductions of those stolen videos and ideas.

52. The unlawful business practices described herein violate the UCL because consumers are likely to be deceived. Defendants knowingly and secretively trained Gemini using unauthorized transcriptions and copies of Plaintiff's 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.

PRAYER FOR RELIEF

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

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

	Case 5:24-cv-04708 D	ocument 1 Filed 08/02/24 Page 11 of 11			
1	g. For injunctive reli	ef as the Court may deem proper; and			
2	h. For an order awarding Plaintiff and the Class their reasonable attorneys' fees and				
3	expenses and costs of suit.				
4	DEMAND FOR TRIAL BY JURY				
5	Pursuant to Federal Rule of Civil Procedure 38(b), Plaintiff demands a trial by jury of any				
6	and all issues in this action so triable of right.				
7					
8	Dated: August 2, 2024	BURSOR & FISHER, P.A.			
9		By: /s/ L. Timothy Fisher			
10		L. Timothy Fisher			
11		L Timothy Fisher (State Bar No. 191626) Joshua B. Glatt (State Bar No. 354064)			
12		1990 North California Blvd., 9th Floor Walnut Creek, CA 94596			
13		Telephone: (925) 300-4455 Facsimile: (925) 407-2700			
14		E-mail: ltfisher@bursor.com			
15		jglatt@bursor.com			
16		BURSOR & FISHER, P.A. Joseph I. Marchese (<i>pro hac vice</i> forthcoming)			
17		Julian C. Diamond (pro hac vice forthcoming)			
18		1330 Avenue of the Americas, 32nd Floor New York, NY 10019			
19		Telephone: (646) 837-7150 Facsimile: (212) 989-9163			
20		E-Mail: jmarchese@bursor.com			
20		jdiamond@bursor.com			
22		Attorneys for Plaintiff			
23					
24					
25 26					
26					
27					
28					
	CLASS ACTION COMPLAINT – JUR	Y TRIAL DEMANDED	10		