

Joseph R. Saveri (State Bar No. 130064)
 Cadio Zirpoli (State Bar No. 179108)
 Travis Manfredi (State Bar No. 281779)
 JOSEPH SAVERI LAW FIRM, LLP
 601 California Street, Suite 1000
 San Francisco, California 94108
 Telephone: (415) 500-6800
 Facsimile: (415) 395-9940
 Email: jsaveri@saverilawfirm.com
 czirpoli@saverilawfirm.com
 tmanfredi@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
 1920 Hillhurst Avenue, #406
 Los Angeles, CA 90027
 Telephone: (323) 968-2632
 Facsimile: (415) 395-9940
 Email: mb@buttericklaw.com

*Counsel for Individual and Representative
 Plaintiffs and the Proposed Class*

**UNITED STATES DISTRICT COURT
 NORTHERN DISTRICT OF CALIFORNIA
 SAN FRANCISCO DIVISION**

J. DOE 1 and J. DOE 2, individually and on
 behalf of all others similarly situated,
 Individual and Representative Plaintiffs,
 v.

GITHUB, INC., a Delaware corporation;
 MICROSOFT CORPORATION, a Washington
 corporation; OPENAI, INC., a Delaware
 nonprofit corporation; OPENAI, L.P., a
 Delaware limited partnership; OPENAI GP,
 L.L.C., a Delaware limited liability company;
 OPENAI STARTUP FUND GP I, L.L.C., a
 Delaware limited liability company; OPENAI
 STARTUP FUND I, L.P., a Delaware limited
 partnership; OPENAI STARTUP FUND
 MANAGEMENT, LLC, a Delaware limited
 liability company,

Defendants.

Case No.

**COMPLAINT
 CLASS ACTION**

DEMAND FOR JURY TRIAL

TABLE OF CONTENTS

I.	OVERVIEW: A BRAVE NEW WORLD OF SOFTWARE PIRACY	1
II.	JURISDICTION AND VENUE.....	4
III.	INTRADISTRICT ASSIGNMENT	4
IV.	PARTIES.....	4
	Plaintiffs	4
	Defendants	5
V.	AGENTS AND CO-CONSPIRATORS	7
VI.	CLASS ALLEGATIONS	8
	A. Class Definitions.....	8
	B. Numerosity.....	9
	C. Typicality.....	9
	D. Commonality & Predominance.....	9
	1. DMCA Violations	10
	2. Contract-Related Conduct	10
	3. Unlawful-Competition Conduct	10
	4. Privacy Violations	10
	5. Injunctive Relief.....	11
	6. Defenses	11
	E. Adequacy	11
	F. Other Class Considerations	11
VII.	FACTUAL ALLEGATIONS	12
	A. Introduction.....	12
	B. Codex Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses	13
	C. Copilot Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses	18
	D. Codex and Copilot Were Trained on Copyrighted Materials Offered Under	

1	E. Copilot Was Launched Despite Its Propensity for Producing Unlawful Outputs	22
2	F. Open-Source Licenses Began to Appear in the Early 1990s	24
3	G. Microsoft Has a History of Flouting Open-Source License Requirements	26
4	H. GitHub Was Designed to Cater to Open-Source Projects	28
5	I. OpenAI Is Intertwined with Microsoft and GitHub.....	30
6	J. Conclusion of Factual Allegations	32
7	VIII. CLAIMS FOR RELIEF.....	33
8	IX. DEMAND FOR JUDGMENT	50
9	X. JURY TRIAL DEMANDED	52
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		

Plaintiffs J. Doe 1 and J. Doe 2 (“Plaintiffs”), on behalf of themselves and all others similarly situated, bring this Class Action Complaint (the “Complaint”) against Defendants GitHub, Inc.; Microsoft Corporation; OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; and OpenAI Startup Fund Management, LLC¹ for violation of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205 (the “DMCA”); violation of the Lanham Act, 15 U.S.C. § 1125; violation of Unfair Competition law, *Cal. Bus. & Prof. Code* §§ 17200, *et seq.*; violation of the California Consumer Privacy Act, *Cal. Civ. Code* § 1798.150 (the “CCPA”); and Breach of Contract regarding the Suggested Licenses, GitHub’s Privacy Statement, and GitHub’s Terms of Service, *Cal. Bus. & Prof. Code* §§ 22575–22579, *Cal. Civ. Code* § 1798.150. Plaintiffs and the Class also bring this Complaint against Defendants for their Tortious Interference in Plaintiffs’ Contractual Relationships; Fraud, and Negligence regarding handling of sensitive data.

I. OVERVIEW: A BRAVE NEW WORLD OF SOFTWARE PIRACY

1. Plaintiffs and the Class are owners of copyright interests in materials made available publicly on GitHub that are subject to various licenses containing conditions for use of those works (the “Licensed Materials.”). All the licenses at issue here (the “Licenses”) contain certain common terms (the “License Terms”).

2. “Artificial Intelligence” is referred to herein as “AI.” AI is defined for the purposes of this Complaint as a computer program that algorithmically simulates human reasoning or inference, often using statistical methods. Machine Learning (“ML”) is a subset of AI in which the behavior of the program is derived from studying a corpus of material called training data.

¹ GitHub, Inc. is referred to as “GitHub.” Microsoft Corporation is referred to as “Microsoft.” OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; and OpenAI Startup Fund Management, LLC are referred to collectively herein as “OpenAI.” Collectively, GitHub, Inc., Microsoft Corporation, OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; and

1 3. GitHub is a company founded in 2008 by a team of open-source enthusiasts. At
2 the time, GitHub's stated goal was to support open-source development, especially by hosting
3 open-source source code on the website github.com. Over the next 10 years, GitHub, based on
4 these representations succeeded wildly, attracting nearly 25 million developers.

5 4. Developers published Licensed Materials on GitHub pursuant to written Licenses.
6 In particular, the most popular ones share a common term: use of the Licensed Materials requires
7 some form of *attribution*, usually by, among other things, including a copy of the license along
8 with the name and copyright notice of the original author.

9 5. On October 26, 2018, Microsoft acquired GitHub for \$7.5 billion. Though some
10 members of the open-source community were skeptical of this union, Microsoft repeated one
11 mantra throughout: "Microsoft Loves Open Source". For the first few years, Microsoft's
12 representations seemed credible.

13 6. Microsoft invested \$1 billion in OpenAI LP in July 2019 at a \$20 billion valuation.
14 In 2020, Microsoft became exclusive licensee of OpenAI's GPT-3 language model—despite
15 OpenAI's continued claims its products are meant to benefit "humanity" at large. In 2021,
16 Microsoft began offering GPT-3 through its Azure cloud-computing platform. On October 20,
17 2022, it was reported that OpenAI "is in advanced talks to raise more funding from Microsoft" at
18 that same \$20 billion valuation. Copilot runs on Microsoft's Azure platform. Microsoft has used
19 Copilot to promote Azure's processing power, particularly regarding AI.

20 7. On information and belief, Microsoft obtained a partial ownership interest in
21 OpenAI in exchange for its \$1 billion investment. As OpenAI's largest investor and largest
22 service provider—specifically in connection with Microsoft's Azure product—Microsoft exerts
23 considerable control over OpenAI.

24 8. In June 2021, GitHub and OpenAI launched Copilot, an AI-based product that
25 promises to assist software coders by providing or filling in blocks of code using AI. GitHub
26 charges Copilot users \$10 per month or \$100 per year for this service. Copilot ignores, violates,
27 and removes the Licenses offered by thousands—possibly millions—of software developers,

28 9. Copilot's training data includes code from GitHub, which is licensed under various open-source licenses. Copilot's output is not derived

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.