

1 Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
2 Christopher K.L. Young (State Bar No. 318371)
Louis A. Kessler (State Bar No. 243703)
3 Elissa A. Buchanan (State Bar No. 249996)
William W. Castillo Guardado (State Bar No. 294159)

4 Holden J. Benon (State Bar No. 325847)
JOSEPH SAVERI LAW FIRM, LLP
5 601 California Street, Suite 1505
San Francisco, California 94108
6 Telephone: (415) 500-6800
Facsimile: (415) 395-9940
7 Email: jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
8 cyoung@saverilawfirm.com
lkessler@saverilawfirm.com
9 eabuchanan@saverilawfirm.com
wcastillo@saverilawfirm.com
10 hbenon@saverilawfirm.com

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

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

16 [Additional counsel on signature page]

17 **UNITED STATES DISTRICT COURT**
18 **NORTHERN DISTRICT OF CALIFORNIA**
OAKLAND DIVISION

19 J. DOE 1 et al.,
20 *Individual and Representative Plaintiffs,*
21 v.
22 GITHUB, INC., et al.,
23 *Defendants.*

Case Nos. 4:22-cv-06823-JST
4:22-cv-07074-JST

JOINT CASE MANAGEMENT STATEMENT

Date: July 12, 2024
Time: 1:30 p.m.
Courtroom: 6, 2nd Floor
Judge: Hon. Jon Tigar

1 Plaintiffs Does 1 through 5 and the proposed class (“Plaintiffs”); Defendants GitHub, Inc.
2 (“GitHub”); Microsoft Corporation (“Microsoft”); and OpenAI, Inc.; OpenAI, LP; OpenAI OpCo, LLC;
3 OpenAI GP, LLC; OpenAI Startup Fund GP I, LLC; OpenAI Startup Fund I, LP; OpenAI Startup Fund
4 Management LLC, OpenAI LLC, OAI Corporation, OpenAI Holdings LLC, OpenAI Holdco LLC,
5 OpenAI Investment LLC, Open AI Startup Fund SPV I, LP, and OpenAI Startup Fund Management, LLC
6 (“OpenAI”) file this Joint Case Management Statement (“CMS”) pursuant to ECF No. 231.

7 **I. STATUS OF DISCOVERY**

8 **A. Plaintiffs’ Discovery Requests**

9 **1. Discovery Propounded by Plaintiffs and Responses from Defendants.**

10 **Requests for Production of Documents (“RFPs”):**

11 **Status of Microsoft and GitHub Document Production:**

12 **Plaintiffs’ Position:** Plaintiffs have attempted to conduct ESI discovery in a manner consistent
13 with Guideline 1.02 (Cooperation) of the Northern District of California’s Guidelines for the Discovery of
14 Electronically Stored Information. As explained below, Plaintiffs’ efforts have been continually frustrated
15 by Defendants’ delays and empty promises. To date, over a year into the litigation, and despite regular
16 meet and confers, few documents have been produced—Microsoft has produced only one document;
17 GitHub has produced roughly one thousand documents, and OpenAI has produced approximately 185
18 documents. The lions-share of the documents that have been produced are publicly available, or are
19 routine discovery that should have been produced months ago (e.g., organizational charts).

20 On July 12, 2023 Plaintiffs served First Sets of Requests for Production on GitHub (49 requests),
21 Microsoft (60 requests), and OpenAI (57 requests). Defendants served responses and objections on August
22 18, 2023. The First Set of Requests was served prior to the Court’s entry of the ESI protocol. In the
23 September 8, 2023 Joint Case Management Statement (ECF No. 153), Defendants attributed their failure
24 to produce any documents to the lack of an ESI Protocol in this case. Then, Defendants ignored repeated
25

1 requests to meet and confer on Plaintiffs' draft ESI Protocol until shortly before the deadline set by this
2 Court. *See* ECF No. 166. The Court entered a version of the Northern District of California ESI Patent
3 Protocol which, among other things, limits email ESI discovery to five custodians and five search terms
4 and does not account for non-email ESI. Even after the entry of an ESI Protocol, document production
5 continued to lag.

6 On March 4, 2024, Plaintiffs served their second set of requests for production to Defendants
7 Microsoft and GitHub. On March 14, 2024, Plaintiffs served their third request for production of
8 documents, emails specifically, pursuant to the ESI protocol, on Defendants Microsoft and GitHub. On
9 May 23, 2024, Plaintiffs sent a Supplemental Third Set of RFPs to Microsoft and GitHub with revised
10 search terms and custodians regarding discovery topics gleaned from the GitHub production.

11 Plaintiffs have sought information regarding corporate structure; identification of those with
12 management or other responsibility for the conduct alleged; the creation, training and operation of the AI
13 models utilized by GitHub Copilot; the source and use of the training data and datasets; use of and
14 compliance with applicable open-source licenses; the AI models trained on Plaintiffs' code outputting
15 matches or near matches of that code; submissions of information to regulators and governmental
16 agencies; public statements regarding these topics; business and financial relationships between and
17 among the parties and their principals, including specifically Microsoft's financial interest in OpenAI's
18 models and Copilot's operation; income, revenue and profit sharing from the sale or distribution of
19 Copilot; and insurance policies and other materials ordinarily produced in cases of this type. Much of this
20 information is classic first-wave discovery common in all cases. Nonetheless, Defendants objected to each
21 and every one of Plaintiffs' RFPs.

22 Microsoft and GitHub first produced a smattering of documents on December 12, 2023. The
23 productions consisted of one Microsoft document—their relevant organizational charts—and 331 GitHub
24 documents. That Microsoft employees were involved in many of these GitHub-sourced conversations
25

1 demonstrates that Microsoft's production of *one* document thus far has been a function of delay and
2 obfuscation, and nothing else. Indeed discovery has shown that Microsoft's employees developed
3 Copilot's duplicate detection feature. That GitHub has offered to make the Copilot model available for on-
4 site inspection is not an excuse for Microsoft's failure to make the most basic document productions
5 regarding relevant information that will provide Plaintiffs and Plaintiffs' experts with the knowledge base
6 to query the model adequately when it is ripe to do so. Nor does inspection serve as a replacement for
7 contemporaneous documents evidencing the development, testing, sourcing or other aspects of Copilot's
8 development or operation. Microsoft has known but failed to disclose that its employees were directly
9 involved in the creation, operation and management of Copilot and its underlying models, and Microsoft
10 employees were directly involved in addressing the problem of Copilot regurgitation its training data.

11 With respect to GitHub specifically, GitHub recently produced 533 documents earlier this month,
12 June 2024. Microsoft has yet to supplement its production, despite repeatedly promising to do so. But
13 Microsoft has recently tried to delay production yet again, citing unspecified technical difficulties related
14 to searching Slack, its internal messaging platform, as reasons they need to delay production and further
15 discuss search terms and custodians. They further have intimated all their productions need to be put on
16 hold for 21 days to assess the proportionality of discovery given the Court's recent order dismissing
17 Plaintiffs DMCA claims. Many, if not all of Plaintiffs' RFPs, however, apply with equal force to the
18 breach of contract claim (and the breach of open-source license claim in particular) as they did to the
19 DMCA claim. Nor do technical difficulties concerning Slack inhibit Microsoft's ability to produce other
20 categories of ESI such as emails. Thus, Microsoft and GitHub's excuses are merely another delay tactic.
21 Such endless delay will eventually require court intervention if not resolved soon and in good faith.

22 Microsoft, in turn, has repeatedly changed its reasons for not producing other go-get documents,
23 from claiming, for example, they are not a party to the contracts and not liable for breaches, to suggesting
24 the contracts are not subject to multiple reasonable interpretations and not ambiguous and can be
25

1 objectively interpreted without discovery into Microsoft's role, to claiming the scope of discovery is
2 contingent on the Motion to Dismiss that was just recently decided. The limited document production
3 provided by GitHub has shown that Microsoft had a role at every stage in the creation and operation of
4 GitHub Copilot, which was a partnership/joint venture between OpenAI, and Microsoft/GitHub. Microsoft
5 employees help test and run the models developed by OpenAI and used by Copilot, and Microsoft runs the
6 operation of Copilot on its Azure cloud computing system. Microsoft employees helped develop the
7 duplication detection feature in Copilot. Microsoft sells Copilot directly through its Visual Studio
8 software. Microsoft has a financial interest in OpenAI and certain of its models, including the models that
9 power Copilot, into which Plaintiffs seek discovery. Plaintiffs await the promised production of the
10 agreements between OpenAI and Microsoft/GitHub regarding Copilot and models that use GitHub
11 repositories as training data that Microsoft agreed to produce.

12 **Microsoft and GitHub's Position:** Plaintiffs are abusing the privilege of the Court's time and
13 attention to this matter with a litany of bickering complaints that neither fairly depict the course of
14 discovery nor raise any disputes ripe for this Court's intervention.

15 From the outset of discovery, GitHub and Microsoft responded that the most pertinent evidence in
16 this case was (1) the source code demonstrating the operation of the product at issue, GitHub Copilot; and
17 (2) data generated by usage of the product. GitHub and Microsoft always recognized that the breach of
18 contract claims would require such discovery, irrespective of the parties' disputes about the remaining
19 scope of the case playing out in multiple motions to dismiss. Rather than focus on what was clearly
20 necessary, Plaintiffs have utterly failed to pursue relevant discovery of these topics efficiently and in good
21 faith. Instead, they focused all of their time and attention on (1) material with at best fringe theories of
22 relevance, and (2) the most burdensome possible approaches to even relevant matters.

23 To wit, as to the first category, source code, in its very first set of document responses in August
24 2023, GitHub tendered its source code for inspection. Both GitHub and Microsoft's discovery responses
25

1 called the source code the “source of truth” about how the GitHub Copilot product operates—a set of facts
2 essential to any claim that the operation of the product breaches a contractual obligation. Counsel for
3 GitHub also explained, repeatedly, that GitHub’s main line of business is to offer systems for managing
4 the development of source code and that GitHub used its own source code management systems to
5 develop its GitHub Copilot product. Accordingly, the vast majority of information about how the product
6 was developed and operates is contained within the GitHub source code repository used to develop
7 GitHub Copilot. Nevertheless, nearly a year later, Plaintiffs *have never set foot in a room with GitHub’s*
8 *source code repository.*

9 Instead, Plaintiffs pursued a burdensome and inefficient companywide fishing expedition for all
10 documents and communications stored in all of the company’s *other* systems while ignoring the nuts and
11 bolts of the product itself. Despite the total inefficiency of Plaintiffs’ approach, GitHub nonetheless
12 collected and produced *its entire Google Docs repository* of documents regarding the development of
13 GitHub Copilot. Plaintiffs misleadingly attempt to characterize this as a paltry production. It was a
14 complete production of exactly what they asked for, without limitation other than privilege. The problem
15 is that what Plaintiffs want is in a place that they apparently don’t want to look. Now, Plaintiffs have
16 moved on from the entire Google Docs repository and are seeking vast quantities of emails and other
17 electronic messages *from more than 30 named Microsoft and GitHub employees and from every member*
18 *of several teams within GitHub and Microsoft* despite the limitations expressly adopted by the Court in the
19 ESI Protocol. Plaintiffs’ approach to discovery about how the product operates is absurd.

20 As to the second category, usage data, Plaintiffs have been equally or even more dilatory when it
21 comes to pursuing the data that would likely be the most pertinent evidence of whether the GitHub Copilot
22 product ever offered any code matching their own. Plaintiffs have no evidence at all that anyone other
23 than a subset of themselves have caused GitHub Copilot to suggest matching code. The obvious place to
24 look for such evidence is in the usage data, which GitHub has been maintaining at its own considerable
25

1 expense and also occasional disruption to its own operations. GitHub first voluntarily disclosed the
2 schema and descriptions of such usage data all the way back in September 2023. Plaintiffs then ignored
3 that disclosure for months. Eventually they asked for a sample of the data, and on February 29, 2024,
4 GitHub voluntarily produced a sample. More months went by without any word from Plaintiffs. Finally,
5 approaching nine months after GitHub’s first disclosure regarding usage data and when the parties were
6 approaching 90 days from the fact discovery cutoff, Plaintiffs finally woke up and provided a draft of a
7 30(b)(6) Deposition Notice designed to assist them in conducting discovery of the usage data. GitHub
8 immediately agreed to produce a witness on many of the proposed topics and has been working to identify
9 the witness and offer dates for this deposition.

10 Regarding ESI, Plaintiffs have been as dilatory and extreme in their approach as on these other
11 issues. On November 16, 2023, Microsoft and GitHub first sent a letter requesting that Plaintiffs serve
12 custodial document requests as outlined in paragraphs 6 and 8 of the ESI Order issued in this matter. *See*
13 ECF No. 168 (“ESI Order”). GitHub specifically identified Slack as a source of custodial data in meet
14 and confers, and also repeatedly said to Plaintiffs’ counsel that the challenges associated with production
15 of Slack data were significant. Plaintiffs waited three months and sent a letter. Finally, on March 14,
16 2024, Plaintiffs propounded custodial document requests. Those requests included requests for both e-
17 mail production and other custodial ESI (like Slack messages) that went far beyond the limits of the ESI
18 Order—as Microsoft and GitHub noted in their objections and responses served on April 15. The parties
19 have conducted *several* meet-and-confers on this topic. Finally, on May 28, 2024, Plaintiffs sent a letter
20 that prioritized custodians and hit counts while continuing to seek ESI from more than 30 named
21 Microsoft and GitHub employees and potentially dozens of additional members of GitHub and Microsoft
22 teams. Microsoft and GitHub are currently in the process of collecting and processing data to provide hit
23 counts for the custodians and search terms that Plaintiffs identified in that letter as their top priority
24 custodians, and offered to assist Plaintiffs with identifying key custodians for specific topics to the extent
25

1 Plaintiffs are unable to determine the proper individuals from Microsoft’s and GitHub’s interrogatory
2 responses and document productions. The volume of Slack data is large and Slack data is singularly
3 resistant to efficient processing. GitHub is continuing to explore means of speeding up the process but it
4 may take another few weeks to be able to process the Slack data and provide the Slack and email hit
5 counts in order to enable the parties to evaluate the proportionality of the search terms.

6 As to the overall scope of relevance of discovery, this case started with 22 claims and now has 2.
7 Microsoft and GitHub have consistently explained that the scope of relevance is tied to the pleadings,
8 indeed one such dispute ripened into a protective order entered by Judge Ryu regarding the six subpoenas
9 and dozens of document requests served by Plaintiffs regarding the GitHub acquisition. As the motion to
10 dismiss the Second Amended Complaint was pending and fully briefed, Microsoft and GitHub offered to
11 supplement their responses to the pending requests for production within 21 days of the order on the
12 motion to dismiss to identify any requests to which they will maintain relevance and
13 burden/proportionality objections in light of the final scope of the pleadings. Microsoft and GitHub
14 reiterated that commitment on June 25, 2024—the day after the Court’s order. Microsoft and GitHub plan
15 to serve those supplemental responses by July 15, 2024, and expect that the parties will thereafter continue
16 their meet-and-confers. Microsoft and GitHub also agreed to produce additional non-custodial documents
17 that are not impacted by the Court order prior to July 15, 2024.

18 There are no disputes ripe for Court intervention.

19 **Status of OpenAI Document Production:**

20 **Plaintiffs’ Position:** OpenAI initially produced 40 documents—almost all publicly available and
21 has since produced 143 more. OpenAI promised in an early April meet and confer to update its
22 interrogatory responses and produce further relevant documents regarding its use of GitHub repositories as
23 training data for the base GPT models, and the quantity of the GitHub-sourced training data used in both
24 the base GPT models and used to “fine tune” the multiple AI models used in Copilot. Copilot utilizes AI
25

1 models that are “fine-tuned” versions of OpenAI’s GPT models, which were trained on the entirety of
2 GitHub public repositories. Plaintiffs await such information to advance the protocol regarding the
3 production, inspection, and/or online access to such training data, and to assist in identifying relevant
4 custodians.

5 Plaintiffs are aware that OpenAI has made a substantial document production numbering tens of
6 thousands of documents in separate matters pending in this district and the Southern District of New York
7 involving ChatGPT (), a generative AI product based on the same GPT models used to develop Codex and
8 Copilot. *See Authors Guild v. OpenAI Inc.*, 1:23-cv-08292, (S.D.N.Y.); *In re OpenAI ChatGPT Litigation*,
9 No. 3:23-cv-03223-AMO. This asymmetry in OpenAI’s document production is puzzling, especially
10 considering that those cases were filed months after this one. Plaintiffs’ counsel is also prosecuting the
11 action pending in this district, and OpenAI is also represented by the same counsel in both the action
12 pending here in the Northern District of California and in the Southern District of New York. Plaintiffs
13 believe many of the documents produced in those actions are relevant here and responsive to Plaintiffs’
14 RFPs. To be clear, Plaintiffs are not asking for “cloned” discovery—Plaintiffs have only asked OpenAI to
15 cross-produce *relevant* documents here because the GPT models at issue there were all trained on
16 GitHub’s public repositories and are used as the “base” models for the code-specific models used by
17 Copilot (including but not limited to Codex, which was a fine-tuned version of GPT-3).

18 Plaintiffs have requested cross-production both orally and in writing. OpenAI has thus far refused
19 to commit to cross-producing any documents. Any burden associated with producing those documents
20 would be minimal given that these documents have already been identified, collected, and produced.
21 Given their relevance and the nominal burden, OpenAI should cross-produce those documents in this
22 matter. Plaintiffs will continue to meet and confer with OpenAI about the relevance and scope of cross-
23 production.

1 **OpenAI’s Position:** Plaintiffs have not served any discovery since they served their First Set of
2 Requests for Production (“RFPs”) in July 2023. OpenAI served its responses and objections in August
3 2023, and has made three productions since that date. These productions comprise non-custodial
4 documents regarding its corporate structure and corporate formation; employee organization charts
5 reflecting those with management or other responsibility for the conduct alleged; documents describing
6 the operation of the relevant technologies; Plaintiffs’ GitHub licenses; and insurance policies applicable to
7 this case. Because OpenAI is a research company that routinely publishes its research, much of the
8 relevant technical documentation Plaintiffs have requested and OpenAI has produced is publicly available.

9 OpenAI is continuing its review of potentially responsive documents and intends to make
10 subsequent rolling productions of relevant documents. This process has been hindered by Plaintiffs’
11 continual focus on demanding modifications to provisions of the final ESI Protocol (applicable only to
12 email documents) rather than moving discovery forward. For example, the ESI Protocol requires that
13 Plaintiffs serve separate requests for email documents. They have not done so. Yet they have repeatedly
14 insisted that the Protocol entered by the Court is unworkable and have aimed to modify it to permit ten
15 email custodians and 15 search terms per email request. But the Court previously considered and rejected
16 Plaintiffs’ arguments for an expanded number of custodians and search terms in the letter brief submitted
17 by the parties prior to entering the Protocol. ECF No. 165 (letter brief filed by parties with ESI proposals
18 for each). When Plaintiffs serve proper email requests under the ESI Protocol that the Court entered,
19 OpenAI will respond to those requests appropriately.

20 Moreover, in an effort to move discovery forward, OpenAI proposed that the parties agree to a
21 mutual exchange of search terms and custodians for non-email document requests. OpenAI shared a
22 proposed list of custodians and search terms with Plaintiffs on May 23, and invited Plaintiffs to request up
23 to ten additional search terms or propose alternate terms or custodians. Plaintiffs did not respond to that
24
25

1 proposal. OpenAI will likely provide a revised list of custodians and search terms in light of the Court’s
2 Order dismissing Plaintiffs’ Section 1202(b) claim. ECF No. 253.

3 While ignoring OpenAI’s efforts to progress with discovery, Plaintiffs sent a letter to OpenAI on
4 June 25, 2024 (1) stating their belief that they are entitled to all documents produced in the *In re OpenAI*
5 *ChatGPT* case and (2) indicating that they believed “[m]any (if not all) of the discovery requests relevant
6 to the DMCA claims and the breach of contract claim are co-extensive” and therefore all propounded
7 discovery requests continue to be relevant.¹ Neither argument has merit.

8 **First**, Plaintiffs point to no specific documents produced in the *In re OpenAI ChatGPT* case that
9 are relevant to the claims here. Instead, they merely claim that the documents produced in the copyright
10 case “appear to be relevant” and demand that “OpenAI produce them given that there is no burden in
11 producing them given that they have already been collected and that proportionality considerations set
12 forth in Rule 26 support the discovery.” Plaintiffs misunderstand the relevant standard—they must make
13 an adequate showing that the requested discovery is relevant to their remaining breach of contract claims.
14 *See, e.g., Ludlow v. Flowers Foods, Inc.*, No. 18-cv-1190, 2019 WL 6252926, at *18 (S.D. Cal. Nov. 22,
15 2019) (requesting “all documents produced in another matter is not generally proper” because the
16 propounding party “cannot meet its burden to establish relevance, as the propounding party is not in a
17 position to even know what they are actually asking for”). Plaintiffs cannot demonstrate the relevance of
18 documents produced in another case where that case involves far different claims, plaintiffs, and OpenAI
19 products. Indeed, this case is the *only* one among various recently filed against OpenAI (both in this
20
21

22
23 ¹ In this joint case management statement, Plaintiffs request (for the first time) that OpenAI produce
24 documents produced in *Authors Guild v. OpenAI Inc.*, 1:23-cv-08292, (S.D.N.Y.) in addition to *In re*
25 *OpenAI ChatGPT Litigation*, No. 3:23-cv-03223-AMO (N.D. Cal.). This request is similarly without
basis.

1 District and others) that involves breach of contract claims for open-source licenses. The other cases (*In*
2 *re OpenAI ChatGPT* included) bring various copyright, trademark, and consumer protection claims.

3 **Second**, Plaintiffs cannot reasonably claim that every discovery request bearing on their dismissed
4 DMCA claim is also relevant to their breach of contract claims. For example, ROG No. 10 asks OpenAI
5 to identify “all Persons [. . .] previously and currently responsible for, or having oversight or control over
6 Ethics Teams relating to artificial intelligence ethics including but not limited to Copilot.” Artificial
7 intelligence ethics are wholly unconnected to the breach of contract claims Plaintiffs bring in their
8 Complaint. Similarly, RFP No. 51 requests “advertising, promotional materials, press releases,
9 magazines, blogs, social media posts or feeds, or websites CONCERNING COPILOT.” These materials
10 have little relevance to the licenses governing Plaintiffs’ code and OpenAI’s purported breach of those
11 licenses. OpenAI will meet and confer with Plaintiffs on these issues but anticipates that Court
12 intervention will be necessary if Plaintiffs do not narrow their current discovery requests to appropriately
13 tailor them to the remaining claims.

14 ***Issues Applicable to All Defendants***

15 **Plaintiffs’ Position:** Over a year after discovery was opened, Microsoft and GitHub have not
16 identified any custodians for document production. OpenAI has only identified four custodians they are
17 searching for non-email ESI; OpenAI has still not identified potential custodians for email ESI. Plaintiffs
18 have been forced to propose custodians for searches based on GitHub’s production and public
19 information. Plaintiffs have been routinely and regularly attempting to negotiate and identify potential
20 custodians throughout the discovery process, including requesting Defendants, consistent with Guideline
21 1.02, to identify who they believe should be potential custodians. However, neither Defendant has
22 identified even the five custodians Plaintiffs are entitled to under the operative ESI Order. Defendants
23 have cited various shifting reasons for their refusal to provide or agree to disclose potential custodians,
24 including the pending motions to dismiss (even though discovery requests have been pending for over a
25

1 year and those requests would be unaffected by the outcomes of the motions to dismiss); and improperly
2 conditioning identification of custodians to other unrelated aspects of discovery, such as demanding
3 Plaintiffs answer premature contention interrogatories before being willing to produce relevant
4 documents. Plaintiffs, for their part, despite having no obligation, consistent with Guideline 1.02's dictate
5 encouraging parties to cooperate in ESI discovery, have met and gone beyond Defendants' conditions, but
6 Defendants continue to refuse to identify custodians.

7 This is complex litigation against three multinational corporations involving two separate types of
8 contract disputes stemming from open-source licenses and GitHub's terms of service, respectively. The
9 terms of the open-source licenses at issue are written broadly and require appropriate discovery into
10 whether Copilot, a commercial generative AI product that was trained on class members' computer code
11 and outputs it, is a "use" or is "based on" that code, such that it triggers the duties under the open-source
12 licenses. Defendants know this—after all, the licenses were attached to Plaintiffs' complaint. GitHub's
13 terms of service and privacy policies are similarly broad: whether selling the Copilot product violates
14 GitHub's promises that it would not and does not "sell" its users' code requires discovery into the
15 objective facts surrounding Microsoft and GitHub's behavior subsequent to Plaintiffs' acceptance of the
16 contracts at issue when they posted their code subject to open-source licenses on GitHub. Further,
17 Plaintiffs seek discovery into the relationship between three multinational corporate defendants who
18 together developed and operate commercial artificial intelligence technology and have complex financial
19 interests with respect to various aspects of Copilot. Contrary to OpenAI's assertion otherwise, marketing
20 of models trained on GitHub repositories that regularly regurgitate their training data is relevant extrinsic
21 evidence to the promise not to sell Plaintiffs' code at issue in this litigation. The parties are continuing to
22 meet and confer on the use of more than five search terms and custodians for email discovery.

23 Plaintiffs have conducted face-to-face meetings with GitHub, Microsoft, and OpenAI on the
24 outstanding discovery requests, as well as countless meet and confers regarding custodians and search
25

1 terms. The parties have continued to discuss the necessary searches and custodians by videoconference
2 and letter and Plaintiffs have arranged for further meetings with Defendants to continue these discussions.
3 Consistent with the principles of cooperation underlying discovery as outlined by the Federal Rules, the
4 Sedona Conference principles, and the Northern District of California’s Guidelines for the Discovery of
5 Electronically Stored Information, Plaintiffs continue to be committed to the principles of cooperation in
6 order to minimize litigation costs and delay; but discovery is a two-way street.

7 Generally, Plaintiffs anticipate Defendants will continue to produce responsive documents through
8 (1) the collection of responsive materials within their possession, custody, or control (*i.e.*, “go and get”)
9 and (2) implementing search terms for custodial ESI as new information is uncovered. Plaintiffs
10 propounded discovery requests months ago to its selection of custodians and search terms which
11 Defendants’ have largely improperly refused to answer or continue to delay in answering. *See In re*
12 *Folding Carton Antitrust Litig.*, 76 F.R.D. 417, 419 (N.D. Ill. 1977) (describing the types of information
13 requested by Plaintiffs here as “classic first-wave discovery”). OpenAI and Microsoft have produced only
14 a fraction of the documents requested despite repeated promises to produce more. Considering
15 Defendants’ continued refusal to comply with their discovery obligations, Plaintiffs anticipate some
16 remaining discovery issues will require court intervention, which will soon be brought to Magistrate Judge
17 Ryu now that the pleadings have been resolved.

18 **OpenAI’s Position:** As discussed in the section above, Plaintiffs have not served *any* requests for
19 email communications. OpenAI has not identified any email custodians because, without knowing what
20 Plaintiffs’ email requests will be, OpenAI cannot determine which custodians would be appropriate. Once
21 Plaintiffs make the appropriate requests, as required by the ESI Protocol, OpenAI will identify appropriate
22 custodians.

23 As to non-email documents, as noted above, OpenAI has already identified four non-email
24 custodians and search terms for non-email documents (despite having no obligation to do so under the ESI
25

1 Protocol, which applies only to email document requests), but has yet to hear back from Plaintiffs on
2 whether they believe additional custodians or search terms would be appropriate. Plaintiffs incorrectly
3 state that they have been “forced to propose Custodians for searches based on GitHub’s production and
4 information.” Plaintiffs have *never* proposed custodians for searches to OpenAI.

5 OpenAI also believes that the scope of continuing discovery should be revisited in light of the
6 Court’s dismissal of the DMCA claim. As discussed above, Plaintiffs have indicated that they not only
7 believe that all pending discovery remains relevant, but also that discovery should be expanded to include
8 all discovery produced in two actions involving copyright claims not at issue here (*Authors Guild and In*
9 *re ChatGPT*). OpenAI believes this position is improper (*see supra*) and anticipates Court intervention
10 will be needed if Plaintiffs refuse to appropriately narrow their discovery to the remaining claims.

11 **Microsoft and GitHub’s Position:** Microsoft and GitHub’s position is outlined above. Microsoft
12 and GitHub are currently in the process of collecting and processing data to provide hit counts for
13 Plaintiffs’ top priority custodians, and have offered to assist Plaintiffs with identifying key custodians for
14 specific topics at their request. There are no disputes ripe for Court intervention.

15 **Third Party Discovery:**

16 **Venture and Banking Subpoena Recipients**

17 **Plaintiffs’ Position:** Beginning in September 2023, Plaintiffs propounded discovery to certain
18 investment firms who were involved in providing investments to GitHub and were involved in the
19 purchase of GitHub by Microsoft. Those requests seek information regarding the open-source licenses at
20 issue in this lawsuit as well other information regarding the transactions. Plaintiffs served the following
21 entities with subpoenas: (1) Morgan Stanley; (2) Hydrazine Capital; (3) Andreesen Horowitz, AH Capital
22 Management LLC; and (4) Sequoia Capital (“VC Recipients”). The VC Recipients confirmed such
23 information is contained in investment decks and other material prepared in connection with analyzing the
24 proposed transactions. Such information is relevant as proof of the elements of the breach of contract
25

1 claim, including Microsoft's intent to use GitHub repositories for training and selling Copilot, damages
2 and other remedies. Microsoft and GitHub intervened, preventing VC Recipients from producing the
3 agreed-upon material and filed a unilateral motion for protective order with Judge Ryu,² who delayed
4 resolution until after Defendants' motions to dismiss the Second Amended Complaint (ECF No. 200) was
5 resolved. *See* ECF No. 226. That order was recently entered. *See* ECF No. 253, dismissing Plaintiffs'
6 DMCA claims. Regardless of the dismissal of the DMCA claims, Defendants' commercial intent and
7 valuation of Plaintiffs' and class members code as training data for Copilot and Microsoft's other
8 commercial generative AI products, as reflected in Microsoft's acquisition of GitHub, is relevant objective
9 evidence of whether Microsoft and GitHub breached GitHub's promise to its users that it does not, and
10 would not, "sell" Plaintiffs' code. Plaintiffs intend to seek the production of these documents forthwith.

11 **Microsoft and GitHub's Position:** As the Court is aware, Plaintiffs' campaign of discovery
12 regarding the GitHub Acquisition (six subpoenas, multiple interrogatories, and dozens of document
13 requests) resulted in a protective order issued by Magistrate Judge Ryu deferring any discovery on this
14 issue until the pleadings were settled. ECF. 226. The GitHub Acquisition has nothing to do with the
15 standard terms of public OSS licenses drafted by third parties, terms of service between GitHub and its
16 customers, whether any GitHub Copilot output matches Plaintiffs' code, whether any failure to attribute
17 such output has caused Plaintiffs harm. The Acquisition simply has nothing to do with any of the
18 elements of the Plaintiffs' claims. This is particular true in light of the Court's dismissal of the unjust
19 enrichment remedy. Plaintiffs sought this discovery to quantify a theory of value to GitHub that is plainly
20 irrelevant. Magistrate Judge Ryu has already provided a mechanism if this issue requires further attention
21 from the Court.

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

There are no disputes ripe for Court intervention.

² Plaintiffs had sought similar material directly from Microsoft and GitHub. Both objected and refused to produce it.

1 **Former Employees of Defendants**

2 **Plaintiffs' Position:** Plaintiffs served subpoenas for production of documents on former OpenAI
3 employees David Willner, Daniela Amodei, and Dario Amodei. Mr. Willner performed agreed searches
4 on his personal communications but has stated what he found is irrelevant and has only to do with broader
5 Trust and Safety issues. Plaintiffs and Mr. Willner agreed to revisit the issue of the relevance of broader
6 Trust and Safety issues of models that used GitHub repositories as training data after the Court's order on
7 Defendants' motion to dismiss. Dario and Daniela Amodei have each agreed to run searches of their
8 personal documents, and discussions regarding the production of their documents is ongoing.

9 **OpenAI's Position:** On March 28, 2024, Plaintiffs served a subpoena on third party David
10 Willner, former Head of Trust & Safety at OpenAI from February 2022 until July 2023.³ The parties have
11 met and conferred numerous times on the propriety of this subpoena given Mr. Willner's knowledge and
12 its relevance to the claims in the case. As repeatedly expressed to Plaintiffs, during Mr. Willner's
13 employment at OpenAI, he was primarily responsible for product safety issues that did *not* encompass
14 copyright, regurgitation, licensing, or any other intellectual property issues. Mr. Willner has no memory
15 of ever working on issues related to copyright or potential regurgitation of training data, a fact also
16 conveyed to Plaintiffs. Despite this lack of relevant knowledge, Plaintiffs refused to withdraw their
17 subpoena. The parties did agree to further discussion of the relevance of the subpoenas after the Court's
18 ruling on OpenAI's motion to dismiss. This meet and confer has not yet occurred, though the dismissal of
19 the DMCA claim eliminates the arguments that Plaintiffs have made to date about why Mr. Willner's
20 testimony or documents would be relevant to the case.

21
22
23
24 _____
25 ³ OpenAI's counsel in this matter was retained to represent Mr. Willner in his individual capacity.

1 **Common Crawl Foundation**

2 **Plaintiffs' Position:** Plaintiffs served a subpoena on Common Crawl Foundation regarding
3 donations from and agreements with the Defendants. Common Crawl runs software that crawls, scrapes,
4 and archives the entire internet, whose corpus was used to train the AI models at issue. Common Crawl
5 refused to produce documents citing disproportionate burden as a non-profit and the unresolved motion to
6 dismiss in the litigation. Plaintiffs intend to pursue such documents now that the motions to dismiss have
7 been resolved and upheld.

8 **Rule 30(b)(6) Deposition Notice on Telemetry Data:**

9 **Plaintiffs' Position:** Plaintiffs are continuing to seek discovery related to Copilot's Telemetry
10 Data (data generated by Copilot that is routinely stored in databases by Defendants). Such data includes
11 the prompt and output pairings of Copilot's use as well as information about Copilot's duplication
12 detection feature, which stores information about the identity of regurgitated training data. Microsoft and
13 GitHub previously suggested seeking relief from storing Telemetry Data as it had previously agreed,
14 through February of 2025. GitHub and Microsoft have indicated they are no longer seeking relief from the
15 data storage agreement and are in the process of identifying a Rule 30(b)(6) deponent(s) and proposing
16 dates for the deposition about the Telemetry Data, including how best to access or produce it for
17 inspection.

18 **Interrogatories:**

19 **Plaintiffs' Position:** On July 19, 2023, Plaintiffs propounded 14 interrogatories each on GitHub,
20 Microsoft, and OpenAI. The interrogatories, among other things, ask Defendants to identify persons
21 responsible for the subject matter or events relevant to Plaintiffs' claims, description or identification of
22 training data, and identification and location of current and former employees who are likely to possess
23 relevant information. These interrogatories are routine first-wave discovery requests propounded in nearly
24 every case. Following an extension, on August 25, 2023, Microsoft provided limited responses to some
25

1 interrogatories, indicating their investigation was not complete, and objected to others. Microsoft, instead
2 of answering interrogatories about its agreements with GitHub and OpenAI regarding Copilot and its
3 underlying AI models, represented that it would produce those agreements. To date, Microsoft has not
4 produced any documents other than one organizational chart, delaying their long-promised initial
5 production of relevant documents. On September 28, 2023, GitHub supplemented its response to
6 Interrogatory 12 regarding contact information of former employees. On December 19, 2023, GitHub
7 submitted a second supplemental response in which it refused to identify the contracts between the parties
8 but stated it would “produce final agreements between Microsoft and GitHub regarding Copilot, and final
9 agreements between Microsoft and GitHub regarding the acquisition of GitHub by Microsoft.” It has yet
10 to do so. It further represented that GitHub did not train or develop the models underlying Copilot or
11 Codex. Plaintiffs continue to discover information showing that Microsoft, GitHub and OpenAI developed
12 numerous models together that are used in Copilot. Plaintiffs are continuing to meet and confer with
13 Defendants regarding Interrogatory responses and supplementation.

14 OpenAI continues to stall in producing the information and supplemental interrogatory responses it
15 promised in April 2024 (but would not provide a date certain to produce) regarding: (1) its use of GitHub
16 repositories as training data for its AI models; (2) which of its models have been utilized as base models in
17 Copilot; and (3) which models it has helped train or “fine-tune” for use in CoPilot. The Parties will
18 continue to negotiate the terms of access to this training data and expect documents concerning the
19 training of its models with GitHub repositories. Discovery regarding the GitHub-sourced training data,
20 including the identification of the sources of the training data, how it was manipulated in “pretraining,” its
21 size and searchability, and its use by Defendants to train the products at issue in this case is a threshold
22 discovery issue and is ongoing. Plaintiffs continue to work with OpenAI in good faith to obtain these
23 discovery materials despite repeated delays.

1 **OpenAI’s Position:** OpenAI has repeatedly explained the burden associated with locating
2 relevant information to respond to several of Plaintiffs’ interrogatories. It has communicated its progress
3 in compiling the requested information during various meet and confers prior to the Court’s Order
4 dismissing the DMCA claim. Following the Order, OpenAI believes many of Plaintiffs’ interrogatories
5 are no longer relevant given the claims remaining in this case. OpenAI intends to meet and confer with
6 Plaintiffs to discuss appropriate tailoring of the requests.

7 Additionally, OpenAI continues to diligently work to allow Plaintiffs access to the data used to
8 train models used for Codex. To that end, OpenAI sent Plaintiffs a draft training data inspection protocol
9 on April 19, and technical details regarding how to effectively inspect that data on May 9. For over two
10 months, Plaintiffs have not provided any edits to the protocol or engaged in further negotiations with
11 OpenAI on possible inspection of the data they claim is central to their case, instead leveling unfounded
12 claims of delay against OpenAI. OpenAI is amenable to continuing to meet and confer with Plaintiffs on
13 access to training data if Plaintiffs can adequately explain its relevance to their remaining breach of
14 contract claims.

15 **Microsoft and GitHub’s Position:** There are no disputes ripe for Court intervention.

16 **B. Defendants’ Discovery Requests**

17 **1. Discovery propounded by Defendants and responses from Plaintiffs.**

18 **Requests for Production of Documents (“RFPs”):**

19 **Plaintiffs’ Position:** On July 12, 2023, OpenAI served a separate set of RFPs on each of the five
20 plaintiffs. On August 11, 2023, all Plaintiffs served objections and responses to OpenAI’s first sets of
21 RFPs and produced responsive documents (131 pages). Plaintiffs’ production is complete. On July 27,
22 2023, GitHub and Microsoft jointly served separate sets of RFPs on each of the five Plaintiffs, broadly
23 seeking material related to threats that Plaintiffs received (the “Joint RFPs”). On August 28, 2023,
24
25

1 Plaintiffs served objections and responses to the Joint RFPs. Plaintiffs' document production is
2 substantially complete.

3 **OpenAI's Position:** OpenAI continues to meet and confer with Plaintiffs regarding Plaintiffs'
4 deficient production, which includes unreadable JSON files, a blank HTML file, emails without any
5 metadata, and improperly redacted PNG files of Slack and other messages. At this juncture, these disputes
6 do not require Court intervention.

7 **Microsoft and GitHub's Position:** Microsoft and GitHub agree with OpenAI that Plaintiffs'
8 production is incomplete and deficient. It will meet and confer with Plaintiffs. At this juncture, these
9 disputes do not require Court intervention.

10 **Interrogatories:**

11 **Plaintiffs' Position:** On July 12, 2023, OpenAI served a separate set of interrogatories on each of
12 the five Plaintiffs. On August 11, 2023, Plaintiffs served objections and complete responses to OpenAI's
13 interrogatories. On October 11, 2023, GitHub propounded interrogatories on the Plaintiffs. All Plaintiffs
14 provided complete responses on November 13, 2023. On January 19, 2024, Plaintiffs served
15 supplemental responses and objections to OpenAI regarding their first set of interrogatories. On March 5,
16 2024, GitHub and Microsoft sent Plaintiffs a letter requesting that Plaintiffs respond more fully to
17 Interrogatories 1-10, and provide largely irrelevant information or information already in their possession.
18 On March 28, 2024, GitHub served their second set of interrogatories on Plaintiffs and Microsoft served
19 their first set. On April 26, 2024, Plaintiffs supplemented their responses to GitHub's first set of
20 interrogatories. On April 29, Plaintiffs responded to Microsoft's first set and GitHub's second set of
21 interrogatories. On May 6, Plaintiff Doe 4 supplemented their response and objections to OpenAI's first
22 set of interrogatories. And on June 6, Plaintiffs supplemented their responses to Microsoft's first set and
23 GitHub's second set of interrogatories. Microsoft and GitHub have not raised any further issues with
24
25

1 Plaintiffs' responses; Plaintiffs believe their responses are substantially complete, but will continue to
2 meet and confer with respect to any perceived deficiencies.

3 **OpenAI's Position:** On December 19, 2023, OpenAI sent Plaintiffs a letter notifying them of
4 deficiencies in their interrogatory responses. After OpenAI and Plaintiffs met and conferred on January 4,
5 2024 regarding these deficiencies, Plaintiffs served supplemental responses on January 19, 2024. On
6 February 8, 2024, OpenAI identified additional deficiencies with Plaintiffs' supplemental responses to
7 Plaintiffs, who agreed to provide a response soon. After substantial delay, Plaintiffs provided additional
8 supplemental responses on May 6, 2024. There are no disputes requiring Court intervention at this time.

9 **Microsoft and GitHub's Position:** On March 5, 2024, Microsoft and GitHub sent a meet and
10 confer letter to Plaintiffs regarding Plaintiffs' deficient responses to GitHub's First Set of Interrogatories.
11 The Parties are still meeting and conferring regarding these responses.

12 On March 28, 2024, Microsoft and GitHub served contention interrogatories seeking, *inter alia*,
13 the facts Plaintiffs contend support each element of their breach of contract claims, whether Plaintiffs
14 contend any provisions of the contracts at issue are ambiguous, and the extrinsic evidence Plaintiffs
15 contend is relevant to the question of interpretation of any allegedly ambiguous provisions. On April 29,
16 2024, Plaintiffs served responses, refusing to respond to some interrogatories and failing to provide facts
17 supporting Plaintiffs contentions in response to others. After meeting and conferring regarding Plaintiffs'
18 deficient responses, on June 06, 2024, Plaintiffs supplemented several of their responses to provide more
19 information regarding their contentions, including that the open-source licenses at issue do not contain
20 ambiguous terms. However, Plaintiffs' supplemental responses failed to respond to several requests in
21 full and improperly incorporated by reference lawyer correspondence and the Second Amended
22 Complaint instead of providing separate and complete responses. Accordingly, on June 14, 2024,
23
24
25

1 Microsoft and GitHub sent another meet and confer letter to Plaintiffs regarding remaining deficiencies.
2 The Parties are still meeting and conferring.

3 There are no disputes ripe for Court intervention.

4 **II. INITIAL DISCLOSURES**

5 Pursuant to Federal Rule of Civil Procedure 26(a)(1)(A), Plaintiffs served their Initial Disclosures
6 on June 30, 2023. Defendants⁴ served their initial disclosures on July 5, 2023. The Parties will supplement
7 Initial Disclosures, when necessary, as required by Rule 26(e). Plaintiffs have recently sought the
8 disclosure of any indemnity agreements between OpenAI and Microsoft that were not previously
9 disclosed. Plaintiffs continue to await production of the indemnity agreements between Microsoft and
10 OpenAI.

11 **III. DISCOVERY REFERRAL**

12 **Plaintiffs' Position:** On November 21, 2023, this Court issued its Order Referring All Discovery
13 Disputes to Magistrate Judge, which directs the parties to bring all future discovery disputes to Magistrate
14 Judge Ryu according to her procedures. ECF No. 175. Pursuant to Judge Ryu's standing order,⁵ the parties
15 must have last met and conferred within five days of bringing any discovery dispute to the Court by joint
16 letter brief. A joint letter brief, ECF No. 213, regarding certain discovery requests related to information
17 about Microsoft's acquisition of GitHub is currently before Judge Ryu, who has postponed ruling on them
18 until after the pleadings have been resolved. ECF No. 226. Another letter brief is before Judge Ryu on the
19 propriety of discovery Plaintiffs seek into: (1) OpenAI's productions to government regulators that relate
20 to LLMs, GitHub, Copilot, or the ownership and control of OpenAI; and (2) persons and entities who own
21 interests of more than 5% of OpenAI. Microsoft's unity of interest, financial interest, control and
22

23 _____
24 ⁴ GitHub, Microsoft, and OpenAI are collectively referred to as "Defendants."

25 ⁵ <https://www.cand.uscourts.gov/dmr-revised-standing-order-072022/>

1 influence over OpenAI are discoverable issues in this matter. A videoconference hearing is scheduled for
2 July 11, 2024, before Judge Ryu.

3 Because Defendants have been uncooperative and reticent in producing relevant responsive
4 documents, Plaintiffs intend to submit a motion or joint stipulation to extend discovery in this matter in
5 the coming weeks.

6 **OpenAI's Position:** OpenAI disagrees that there is any need to extend discovery at this time,
7 especially considering the narrowed scope of discovery following dismissal of Plaintiffs' DMCA claim.
8 Any purported delay in discovery has been the result of Plaintiffs' refusal to engage with OpenAI to move
9 the discovery process forward. Nevertheless, OpenAI is willing to consider a reasonable extension of the
10 case schedule, should Plaintiffs propose one.

11 **Microsoft and GitHub's Position:** See above.

12 **IV. CASE MANAGEMENT ORDERS**

13 On August 7, 2023, the Court entered an Amended Stipulated Protective Order For Litigation
14 Involving Patents, Highly Sensitive Confidential Information AND/OR Trade Secrets. ECF No. 145
15 ("Protective Order"). On October 30, 2023, the Court entered a Stipulated Order Concerning Expert
16 Discovery. ECF No. 167 ("Expert Discovery Order"). On November 1, 2023, the Court entered an Order
17 Re Discovery of Electronically Stored Information. ECF No. 168 ("ESI Protocol"). On November 21,
18 2023, the Court ordered all discovery disputes be referred to Magistrate Judge Donna M. Ryu. ECF No.
19 175. On December 12, 2023, the Court ordered the case management conference scheduled for December
20 20, 2023, to be continued to March 22, 2024. ECF No. 186.

21 **V. ISSUES REQUIRING THE COURT'S ATTENTION**

22 **A. Scope of Discovery**

23 **Plaintiffs' Position:** The Plaintiffs anticipate that the Court or Magistrate Judge Ryu will be asked
24 to determine the continued applicability of the model patent ESI protocol to this case if the parties cannot
25

1 agree on the appropriate number of custodians and search terms to satisfy the requisite diligence with
2 which reasonable searches must be conducted under Fed. R. Civ. P. 34. Due to the significant delay in
3 document production and the dearth of documents produced, the model patent order has proven
4 unworkable. As discussed above, this matter is not typical patent litigation and involves far more nuanced
5 issues regarding Defendants’ creation of Copilot and Defendants’ knowledge of and intent to breach open-
6 source licenses and GitHub’s terms and conditions wherein it promises not to sell its users’ code.

7 Further, now that the DMCA claims have been dismissed, Defendants are already attempting to
8 cancel meet and confers to supposedly “assess what discovery remains relevant and necessary.” Given that
9 the open-source licenses are licenses whose duties are triggered by use of the licensed code, the scope of
10 discovery will be largely the same regardless of the disposition of the DMCA claims. Discovery into how
11 Copilot copied Plaintiffs’ code for training and creates infringing works “based on” or “using” Plaintiffs’
12 code is relevant to whether the training and operation of Copilot triggers the duties of the open-source
13 licenses, and whether Defendants breached those duties. Further, extrinsic evidence is discoverable and
14 admissible to establish the meaning of the broad and ambiguous language of the contracts at issue, as they
15 are applied to the training and sale of Copilot. This includes objective facts that inform the meaning of
16 contract terms and the actions of the parties subsequent to the contract creation—when Plaintiffs and class
17 members posted their code on GitHub under an open-source license.

18 Plaintiffs anticipate the potential need for court intervention to clarify the scope of discovery
19 regarding the breach of the open-source licenses and GitHub’s policies and promises not to sell its users’
20 code.

21 **OpenAI’s Position:** OpenAI believes the operative ESI Protocol is appropriately tailored to
22 discovery in this case, particularly after dismissal of the DMCA claim. Plaintiffs have continually
23 attempted to re-negotiate the provisions of the Protocol, raising the same arguments already considered—
24 and rejected—by this Court. *See* ECF No. 165 (arguing without basis that this case “is much more
25

1 complex than non-class patent or trade secret matters,” necessitating 10 custodians and other provisions
2 rejected by the Court).

3 Contrary to Plaintiffs’ assertions, they have made no attempt to contact OpenAI to schedule a meet
4 and confer after dismissal of the DMCA claim. OpenAI has made no attempt to cancel any meet and
5 confers and remains willing to discuss the scope of discovery given the Court’s narrowing of the claims in
6 this case. As mentioned above, OpenAI anticipates the potential need for Court intervention should
7 Plaintiffs refuse to appropriately narrow their discovery requests to the claims remaining in this case.

8 **Microsoft and GitHub’s Position:** Plaintiffs’ approach to the ESI Order has been to delay and
9 then completely ignore its limitations on the number of custodians and search terms. This Court made
10 clear that it would pursue a baseball arbitration approach to the ESI Order. As a result, Defendants
11 proposed a reasonable order based on the form used by the Court. Plaintiffs proposed a 46-page
12 monstrosity that bore little resemblance to the model order, and bore no relationship to the type of case at
13 issue. Unsurprisingly, the Court picked Defendants’ form of order. Ever since then, Plaintiffs have been
14 acting as if the Court chose their form of order instead. They refused to serve custodial discovery
15 requests. After they finally served custodial discovery requests, those requests specified far more
16 custodians and search terms than authorized. Then they took the view that the Order authorized an
17 unlimited number of custodians and search terms, all they had to do was serve as many requests as they
18 wanted and specify different custodians and terms for every separate request. Plaintiffs kept looking for
19 an excuse to declare “impasse” and try to seek the Court’s reconsideration of the ESI Protocol. Despite
20 this absurd position, Microsoft and GitHub kept trying to move the ball forward and keep focused on the
21 actual relevance and proportionality of the custodial discovery requests. Those efforts ultimately resulted
22 in the May 28, 2024 letter described above where Plaintiffs prioritized their requests among a list of
23 dozens of custodians, and Microsoft and GitHub are processing data and running hit counts on the priority
24 custodians.

1 Plaintiffs' statement that Defendants are "attempting to cancel meet and confers" is patently false.
2 *Prior to* the Court's order on the motion to dismiss, Microsoft and GitHub informed Plaintiffs that they
3 would supplement their discovery responses to identify any requests to which they will maintain relevance
4 and burden/proportionality objections in light of the final scope of the pleadings. The day after the
5 Court's order, in response to a request for a meet-and-confer on non-custodial documents, Microsoft and
6 GitHub reiterated their commitment to providing that supplementation within 21 days.

7 **VI. ISSUES BEARING ON THE PROGRESS OF THE CASE**

8 **A. Interlocutory Appeal**

9 **Plaintiffs' Position:** Plaintiffs are considering appealing this Court's order dismissing Plaintiffs'
10 DMCA § 1202 and unjust enrichment claims with prejudice in its recent order on the Defendant's motion
11 to dismiss the Second Amended Complaint pursuant to 28 U.S.C. § 1292(b). ECF No. 253 at 15. Should
12 Plaintiffs move the Court to certify its order for interlocutory appeal under 28 U.S.C. § 1292(b), Plaintiffs
13 do not believe litigation must be stayed during this appeal and that their prosecution of the remaining
14 contract claims can continue. Plaintiffs believe the factors for granting such leave exists here: (1) that
15 there be a controlling question of law, (2) that there be substantial grounds for difference of opinion, and
16 (3) that an immediate appeal may materially advance the ultimate termination of the litigation. *Id.* An
17 immediate appeal would advance the issue of the applicability of the DMCA to generative AI products
18 that are trained on copyrighted material and output portions of code. It would further advance the issue of
19 appropriate remedies for breach of promises not to sell code where no remedy for doing so was
20 contemplated or expressed in the contract.

21 **OpenAI's Position:** OpenAI will oppose any motion to certify interlocutory appeal of the Court's
22 recent order. The Court's order (1) does not present a controlling question of law, as piecemeal appeals
23 on the DMCA dismissal would not have any immediate effect on the course of the litigation; (2) Plaintiffs
24 do not identify (and there does not exist) a substantial ground for difference of opinion on the legal issues
25

1 on which this Court ruled in the motion to dismiss; and (3) the requested appeal would not “materially
2 advance” the termination of litigation as it would only implicate a portion of Plaintiffs’ class-wide claims
3 and would not encourage a quick end to the lawsuit. OpenAI looks forward to more fully developing
4 these points in the event Plaintiffs choose to file a motion to which they would be responsive.

5 **Microsoft and GitHub’s Position:** An interlocutory appeal is not appropriate. The Court’s
6 decision does not involve the kind of “exceptional circumstances” that would warrant a departure from the
7 normal process of awaiting final judgment for appellate review. *See Stiner v. Brookdale Senior Living,*
8 *Inc.*, 383 F. Supp. 3d 949, 957 (N.D. Cal. 2019). Among other things, there are not “substantial grounds
9 for difference of opinion” on the Court’s rulings because this Court’s decisions have been in line with all
10 others in this Circuit. *Id.* As to both the DMCA § 1202 and unjust enrichment claims, courts in this
11 Circuit are in agreement. Claims of occasional out-of-circuit district court decisions following different
12 analyses are not enough to create a substantial grounds for difference. *See Couch v. Telescope Inc.*, 611
13 F.3d 629, 633 (9th Cir. 2010) (“Courts traditionally will find that a substantial ground for difference of
14 opinion exists where ‘the circuits are in dispute on the question and the court of appeals of the circuit has
15 not spoken on the point’”) (quoting 3 Federal Procedure, Lawyers Edition § 3:212 (2010) (emphasis
16 added)). And interlocutory review would serve little purpose with respect to unjust enrichment, because
17 only the California Supreme Court can definitively interpret California law. *See Weiner v. San Diego*
18 *Cnty.*, 210 F.3d 1025, 1028-29 (9th Cir. 2000). Nor would interlocutory review “materially advance the
19 ultimate termination of the litigation.” 28 U.S.C. § 1292(b). As to both claims, an immediate appeal
20 would not further the ultimate termination of the litigation but instead prolong it.

1 Dated: July 5, 2024

By: /s/ Joseph R. Saveri
 Joseph R. Saveri

3 Joseph R. Saveri (State Bar No. 130064)
4 Cadio Zirpoli (State Bar No. 179108)
5 Christopher K.L. Young (State Bar No. 318371)
6 Louis A. Kessler (State Bar No. 243703)
7 Elissa A. Buchanan (State Bar No. 249996)
8 William W. Castillo Guardado (State Bar No. 294159)
9 Holden J. Benon (State Bar No. 325847)
10 **JOSEPH SAVERI LAW FIRM, LLP**
11 601 California Street, Suite 1505
12 San Francisco, California 94108
13 Telephone: (415) 500-6800
14 Facsimile: (415) 395-9940
15 Email: jsaveri@saverilawfirm.com
16 czirpoli@saverilawfirm.com
17 cyoung@saverilawfirm.com
18 lkessler@saverilawfirm.com
19 eabuchanan@saverilawfirm.com
20 wcastillo@saverilawfirm.com
21 hbenon@saverilawfirm.com

22 Matthew Butterick (State Bar No. 250953)
23 1920 Hillhurst Avenue, #406
24 Los Angeles, CA 90027
25 Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@buttericklaw.com

Counsel for Plaintiffs and the Proposed Class

1 Dated: July 5, 2024

By: /s/ Annette L. Hurst
Annette L. Hurst

Annette L. Hurst (State Bar No. 148738)
ORRICK, HERRINGTON & SUTCLIFFE LLP
The Orrick Building
405 Howard Street
San Francisco, CA 94105
Telephone: 415-773-5700
Facsimile: 415-773-5759
Email: ahurst@orrick.com

Counsel for GitHub, Inc. and Microsoft Corporation

8 Dated: July 5, 2024

9 By: /s/ Joseph C. Gratz
Joseph C. Gratz

10 Joseph C. Gratz (State Bar No. 240676)
MORRISON & FOERSTER LLP
425 Market Street
San Francisco, CA 94105
Telephone: 415-268-7000
Facsimile: 415-268-7522
Email: jgratz@mofocom

14 *Attorneys for Defendants OPENAI, INC., OPENAI,*
15 *L.P., OPENAI OPCO, L.L.C., OPENAI GP, L.L.C.,*
16 *OPENAI STARTUP FUND GP I, L.L.C., OPENAI*
17 *STARTUP FUND I, L.P., OPENAI STARTUP FUND*
18 *MANAGEMENT, LLC, OPENAI, L.L.C., OPENAI*
19 *GLOBAL, LLC, OAI CORPORATION, OPENAI*
20 *HOLDINGS, LLC, OPENAI HOLDCO, LLC, OPENAI*
21 *INVESTMENT LLC, OPENAI STARTUP FUND SPV I,*
22 *L.P., and OPENAI STARTUP FUND SPV GP I, L.L.C.,*

L.R. 5-1 SIGNATURE ATTESTATION

As the ECF user whose user ID and password are utilized in the filing of this document, I attest under penalty of perjury that concurrence in the filing of the document has been obtained from each of the other signatories.

Dated: July 5, 2024

/s/ Joseph R. Saveri

Joseph R. Saveri