October 3, 2025

Via Electronic Filing

The Honorable Susan van Keulen
United States District Court for the Northern District of California
San Jose Courthouse, Courtroom 6 – 4th Floor
280 South First Street
San Jose, CA 95113

Re: ***In re Google Generative AI Copyright Litigation***
Master File Case No. 5:23-cv-03440-EKL-SVK
Consolidated Case No. 5:24-cv-02531-EKL-SVK

Pursuant to Section 7(b) of the Court's Civil and Discovery Referral Matters Standing Order, Plaintiffs and Defendant Google LLC ("Google") respectfully submit this joint statement regarding production and confirmation of use of certain datasets. Fact discovery closes in 133 days, on February 13, 2026. Counsel have met and conferred but remain at impasse.

Introduction. Through testimony taken on September 12, Plaintiffs have discovered that Google acquired and ingested for training and building the PaLM, GlaM, LaMDA, Bard, Gemini, and Imagen models ("Models"), data corpora [redacted], FineWeb, FineWeb2, FineWeb-edu, and Common Crawl (The "Ingested Datasets"). These datasets, which are derived from web crawled sources, likely contain copyrighted class works, including works acquired from pirated sources. *See infra,* at 3-4. Plaintiffs allege that Google's unauthorized copying and use of copyrighted works to build and train the Models is copyright infringement. *See* ECF 234 ¶¶ 1, 2, 6, 11, 109, 119, 128, 159, 169. Plaintiffs' theory includes the ingestion process as the first step in the infringement. Plaintiffs are entitled to challenge (and obtain discovery as to) the whole chain of infringement from the entire training process, not just the pre-training step as Google would have it. If Google acquired and placed these corpora into its training pipeline, *i.e.*, used them for pre-training, mid-training, post-training (fine-tuning), and ablation studies (testing how training data effects model performance), then these corpora were used to build and train the Models. This discovery is thus tethered directly to Plaintiffs' allegations. *Id.* Plaintiffs asked Google to make these data corpora available in the training data review platform, but Google refuses based on Google's view that this case is only about pre-training, and not the entire training pipeline. Google also asserts, with regard to [redacted], that Plaintiffs should be constrained only to prior training data selections based on a limited Google-controlled record and should not be informed by later produced evidence. This is nonsense. Plaintiffs ask that the Court compel Google to produce the Ingested Datasets. Further, Plaintiffs ask Google to reveal whether it ingested The Pile, RedPajama, and RedPajama2, as Google's witness could—or would—not.[1]

Google has previously produced exemplar datasets containing purloined copies of Plaintiffs' and class members' works. After propounding discovery requests in January 2025 and engaging in exhaustive conferrals throughout the spring, Google agreed to produce a limited set

[1] To put to rest any contention from Google that the brief is not double-spaced in compliance with the Court's standing order, that is false. *See Sameer v. Khera*, 2018 WL 3472557, at *1 (E.D. Cal. July 18, 2018) ("double-spaced mean[s], at minimum, 24-point line spacing"); *Andrich v. Glynn*, 2023 WL 4847592, at *3 (D. Ariz. July 28, 2023).

of data cards that Plaintiffs were told to use to select a limited number of exemplar datasets for prioritized production. In May, Plaintiffs identified over twenty exemplar datasets for prioritized production, which the Court ordered Google to make available by June 23, 2025. *See* ECF 155 at 2. Substantial delays caused by Google's technological infrastructure prevented access until July 3 with subsequent issues further delaying Plaintiffs' access. *See* ECF 222 at 6; ECF 231 at 2.

This Court has recognized the centrality of training data, and ordered Google to make available datasets used in training Google's models. *See* Hr'g Tr. of June 18, 2025; ECF 155 at 2. These newly identified datasets are relevant for the same reasons that necessitated production of the previously provided exemplar datasets: they contain copyrighted class works and are common evidence of Google's knowing ingestion and use of copyrighted works, including pirated works, to build and train its commercial generative AI models.

Plaintiffs' request is timely. Plaintiffs were not made aware that Google had acquired, ingested, and used for training these datasets, nor did Plaintiffs learn they contained copyrighted material until the deposition of Dr. Xiao on September 12. Plaintiffs could not have previously identified them for production. Plaintiffs promptly requested production of these datasets one business day later. Google has not responded to Plaintiffs' several requests for these datasets. Google's silence amounts to a refusal to produce these highly relevant datasets. Google's refusal to produce is an attempt to shield core evidence from Plaintiffs. The datasets are central to Plaintiffs' claims that Google infringed copyrights to build the Models.

Factual Background. Plaintiffs identified priority datasets in May based on a review of data cards produced by Google. It has recently become clear that the data cards revealed only a portion of the data Google used for training and building its models. As Google's tardy custodial productions trickled in towards the latter half of summer, the documents hinted at other potential sources of data Google considered for ingestion and training. Once Plaintiffs confirmed that Google acquired, downloaded, and otherwise ingested these data sources, and in some cases, confirmed their use as training data, Plaintiffs promptly requested their production.

Plaintiffs confirmed that Google ingested and used these datasets through recent testimony by Dr. Xiao. He testified that the Core Data Acquisition team ("CDA") ingested the Ingested Datasets. Ex. 2, Xiao Dep., 146:15-150:3; 231:10-18; 313:19-314:1. He further testified that these datasets were made available to Google DeepMind for use as training material. *Id.*; *see also* GOOG-AIC-000374443 at -456.

Dr. Xiao testified that after Gemini 1, the CDA was tasked with creating [redacted]. Ex. 2, 308:8-14; 311:21-312:4. This dataset was even larger and more expansive than [redacted]. *Id.*, 69:13-20. What is more, Dr. Xiao admitted that Google also targeted [redacted], which are core class works. *Id.*, 53:17-54:1; 55:10-56:12; 57:8-18; 58:20-59:4; 332:16-334:9; *see also* GOOG-AIC-000374306 [redacted]

FineWeb is "a 15-trillion token dataset derived from 96 Common Crawl snapshots that produces better-performing LLMs than other open pretraining datasets." Guillhereme Penedo et al., *The FineWeb Datasets: Decanting the Web for the Finest Text Data at Scale* (Oct. 31, 2024), https://arxiv.org/abs/2406.17557. FineWeb has been called a new version of "The Pile," another open pretraining dataset known to include pirated material (including Books3). *See* Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, (Dec. 30, 2020), https://arxiv.org/abs/2101.00027 ("We introduce a new filtered subset of Common Crawl, Pile-CC, with improved extraction quality"); *see also* Alex Reisner, *Revealed: The Authors Whose Pirated Books Are Powering Generative AI*, The Atlantic, (Aug. 19, 2023) (reporting on The Pile's distribution and takedown). Dr. Xiao testified that Google ingested and downloaded these datasets for use as AI training material. Ex. 2, 146:15-150:3. That is more than speculation. It is an admission that these corpora entered Google's training pipeline. Given the likelihood that these datasets contain copyrighted and pirated works, they are relevant to class issues. *See Bartz v. Anthropic PBC*, 2025 WL 1993577, at *14 (N.D. Cal. July 17, 2025) (finding defendant "liable for all pirated copies regardless of whether they were used for training…").

Dr. Xiao was also asked at his deposition about other datasets known to be used as training data and to include pirated material: The Pile; RedPajama; RedPajama2. Ex. 2, 148:12-14; 229: 5-16. He testified that the CDA did not ingest those corpora but could not confirm whether Google may have used these datasets (or portions thereof) in building or training the Models. *Id.* These datasets are known to contain pirated copyrighted works. *See* Maurice Weber et al., *RedPajama: An Open Dataset for Training Large Language Models*, (Nov. 19, 2024), https://arxiv.org/abs/2411.12372; *see also* Alex Reisner, *Revealed: The Authors Whose Pirated Books Are Powering Generative AI*, The Atlantic, (Aug. 19, 2023). Plaintiffs are entitled to know whether these datasets were used to develop or train the Models, and if so, their production.

The identified datasets are central to the case and relevant to class certification. The production of datasets used to train, improve, or evaluate the Models is highly relevant to central issues in this case. Common evidence from these datasets will show precisely which class works are at issue. This is a fact. Even Google acknowledges that training data is central class discovery. *See* Hr'g Tr. of June 18, 2025. Production of these datasets is fully consistent with Rule 26(b)(1), which provides that "[p]arties may obtain discovery regarding **any** nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case." Fed. R. Civ. P. 26(b)(1) (emphasis added). Further, these datasets are likely to contain copyrighted class works, as well as pirated works, which alone establishes their relevance. *See Bartz*, 2025 WL 1993577, at *18. Class certification requires a "rigorous" analysis into whether common questions predominate, and class treatment is manageable and administratively feasible. *See Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 359 (2011). Here, the composition and handling, including actual ingestion, of training data are core to predominance. *See Owino v. CoreCivic, Inc.*, 60 F. 4th 437, 445 (9th Cir. 2022); *see also Comcast Corp. v. Behrend*, 569 U.S. 27, 34-35 (2013).

The datasets requested are relevant to liability and common proof. The requested datasets themselves provide a common basis to establish Google's liability. *In re Napster, Inc. Copyright Litig.*, 2005 WL 1287611, at *7 (N.D. Cal. June 1, 2005) (class adjudication appropriate "when common proof establishing liability is a handful of discrete datasets"). Indeed, it is the best

proof of whether the data Google downloaded and ingested for use as training data contained unlawfully acquired copyrighted data. *Id.*

The datasets are relevant to willfulness and statutory damages. The copyright statute provides enhanced statutory damages in cases of willful infringement. 17 U.S.C. § 504(c)(2). Willful infringement occurs when "defendant recklessly disregarded the possibility that its conduct represented infringement" or acted "with knowledge that [its] conduct constitutes copyright infringement." *Wall Mountain Co. v. Edwards*, 2010 WL 4940778, at *2 (N.D. Cal. Nov. 30, 2010). Google's acquisition, ingestion, development, and training on datasets containing copyrighted material is squarely applicable to classwide proof of willful infringement. *See Bartz*, 2025 WL 1993577, at *14. ("[T]o determine a statutory award, a jury must first make findings as to the infringer's mental state…"). Further, because FineWeb was publicly touted as a new version of "The Pile," similarly sourced from Common Crawl, Google's knowing acquisition of a dataset compiled from the same source as a dataset containing pirated works supports willfulness.

The datasets are relevant to classwide measurement of damages. The composition of the datasets will enable classwide identification of works and calculation of aggregate damages or other remedies. *In re Napster, Inc. Copyright Litig.*, 2005 WL 1287611, at *7.

Production of the datasets is not burdensome. Production of the requested datasets will increase efficiency and impose minimal burden on Google. Plaintiffs have only selected a few of the most relevant newly-identified datasets for inspection. *See Woodard v. Labrada*, 2017 WL 10702139, at *4 (C.D. Cal. Apr. 19, 2017). Each Rule26(b)(1) factor favors production. The issues are of substantial public importance given Google's own claims about Generative AI's economic impact and the breadth of the alleged infringement. The amount in controversy is in the billions. Google has the resources to locate and produce the training datasets under its exclusive control.

Plaintiffs move to compel production of the Ingested Datasets. Ex. 2, 146:15-150:3. With respect to The Pile, RedPajama, RedPajama2, Plaintiffs seek confirmation whether Google used any portion of these datasets, to build or train any of the Models, and, if so, production thereof.

Google's Position: This is yet another sideshow. Plaintiffs reach beyond the four corners of the operative complaint (which Judge Lee has foreclosed from further amendment) by making an untimely demand for additional, massive datasets, all but one of which were never even used to train models at issue in the case. Their demand would impose substantial undue burden and flout the Court-endorsed, negotiated framework for training-data discovery.[2]

To begin, Plaintiffs seek seven datasets (FineWeb, FineWeb2, FineWeb-edu, Common Crawl, The Pile, RedPajama, and RedPajama2) that are outside the scope of this case because they were never used to train any models at issue. Plaintiffs' operative complaint limits this case to works actually used to train Google's relevant generative AI models, and Plaintiffs are bound by those allegations. The class they pleaded consists of "all persons … who owned a [U.S. copyright] in any work ***used by Google to train*** Google's Generative AI Models during the Class Period." ECF No. 234 ¶ 163 (emphasis added); *see also* ECF No. 128 at 5 (Judge Lee's order on motion to strike requiring as an "objective criteria" that a work "was used by Google to train" its models). Plaintiffs' allegations likewise tie alleged infringement to use in training. *See, e.g.*, ECF No. 234 ¶ 2 (alleging copies are made "during model training"); *id.* ¶¶ 27, 64 (alleging works were "copied and reproduced … to train Gemini and Imagen"). Moreover, Judge Lee has limited the "Generative AI Models" at issue in this case to a specific set—PaLM, GLam, LaMDA, Bard, Gemini, and Imagen—and denied Plaintiffs leave to add others. ECF No. 216 at 10, 22.

Plaintiffs concede they have ***no evidence*** that these seven datasets were used to train any at-issue model, stretching a lone witness' testimony to find even a whisper of support for that key proposition.[3] Instead, they pivot to a new theory: Google "acquired, downloaded, and otherwise ingested these data sources"—treating "ingestion" as a separate, actionable wrong divorced from

[2] For this brief (unlike prior ones), Plaintiffs insisted on 24-point spacing rather than double so they could add more words and lines. If the Court has no objection, neither does Google.

[3] Plaintiffs cite Dr. Li Xiao's testimony and argue that these datasets "were made available to Google DeepMind for use as training material," *supra* at 3, but availability is not training use. More importantly, the deponent, who testified in his individual capacity and does not work at Deepmind, did not testify that these datasets were used as training material and, in fact, repeatedly disclaimed personal knowledge of what DeepMind actually used to train. *See, e.g.*, Ex. 3, Xiao Dep. Tr. at 81:3-5; 87:21-88:3; 151:1-7.

training use. *Supra* at 2.[4] That new theory lies outside Plaintiffs' case.[5] Nor is the theory salvaged through speculation that Google "placed these corpora into its training pipeline." *Supra* at 1 ("***If*** Google … placed these corpora into its training pipeline, … ***then*** [they] were used to build and train the Models." (emphases added)). Plaintiffs provide ***no evidence*** that the datasets entered any such "pipeline"—they merely assert it conditionally and hypothetically. In any event, the complaint does not refer to a training ***pipeline***; it alleges liability only for actual training ***use***.

Three weeks ago, Judge Lee dismissed certain claims ***without leave to amend***, explaining that Plaintiffs have repeatedly "failed to cure pleading deficiencies" despite multiple opportunities and that amendment would be "futile" and prejudicial. ECF No. 216 at 20, 22. Plaintiffs cannot end run by discovery what the Court firmly foreclosed by Order. *See, e.g.*, *Hiramanek v. Clark*, 2016 WL 217255, at *1 (N.D. Cal. Jan. 19, 2016) (denying discovery motions "relitigat[ing] issues on which the court has already ruled"). Having chosen to define their class around works used for training, there is no cause for discovery of datasets that were not used to train.

Plaintiffs' counsel tried this same gambit in *Nazemian v. Nvidia*, demanding datasets beyond those allegedly used to train. Judge Tigar affirmed Magistrate Judge Kim's order limiting discovery to the sole dataset plaintiffs plausibly alleged was used in training and denied discovery into other datasets, finding no legal error or abuse of discretion in "refusing to grant discovery … on theories of infringement that exceed the scope of the allegations in the complaint," and noting that "[c]ourts in this district have routinely denied discovery regarding theories that exceed the scope of the operative complaint." 2025 U.S. Dist. LEXIS 189507, at *9 (N.D. Cal. Sept. 25, 2025); *see also Andersen v. Stability AI Ltd.*, 23-cv-00201-WHO (N.D. Cal. June 27, 2025), ECF No. 314 at 2 ("Plaintiffs cannot use discovery 'as a fishing expedition for evidence of claims that

[4] Indeed, Plaintiffs cannot even allege that Google "ingested" The Pile, RedPajama, and RedPajama-2, instead asking in their brief that Google confirm "whether it ingested" those sets, *supra* at 1, despite having no outstanding discovery request to that effect.

[5] Plaintiffs' brief invokes "ingest" (or a variant) 14 times—treating downloading or collection as a standalone wrong untethered to training. The complaint, however, uses the term only twice—and only to describe ingestion of data ***by models during training***, not mere data acquisition. *See* ECF No. 234 ¶ 116 (data "ingested by the model during training"); *id.* ¶ 117 (model image generation "imitat[es] the protected expression ingested from the training dataset").

have not been properly pled' or potential future lawsuits." (citation omitted)); *In re Mosaic LLM Litig.*, 2025 WL 2402677, at *3 (N.D. Cal. Aug. 19, 2025) ("Discovery is not a 'fishing expedition.' Rather, the Federal Rules limit discovery to material that is relevant and proportional to claims that survive past the pleading stage.") (citation omitted); *Nazemian v. Nvidia Corp.*, No. 24-cv-01454-JST (SK), ECF No. 185 at 2 (N.D. Cal. Oct. 2, 2025) (refusing (again) "to expand the universe of models for production" beyond the pleadings). As Judge Tigar explained, the discovery limits simply reflected that "the limitations on the proposed class stem from Plaintiffs' own operative complaint." *Nazemian*, 2025 U.S. Dist. LEXIS 189507, at *10. So too here.

Plaintiffs' reliance on *Bartz v. Anthropic* is misplaced. There, Judge Alsup considered a broader infringement theory, that a generative AI service was allegedly "liable for all pirated copies [of works] regardless of whether they were used for training." 2025 WL 1993577, at *14 (N.D. Cal. July 17, 2025). The *Bartz* plaintiffs claimed their works were infringed not just through use from "LLM training," but also through "research, or development." First Am. Class Action Compl., Case No. 3:24-cv-05417-WHA (N.D. Cal. Dec. 4, 2024), ECF No. 70 ¶ 63. Given that expressly broader case scope, Judge Alsup allowed discovery into books housed in a "general 'research library'" without regard to whether the contents were ultimately used to train. *Bartz v. Anthropic PBC*, __ F. Supp. 3d __, 2025 WL 1741691, at *3 (N.D. Cal. June 23, 2025). By contrast, Plaintiffs here limited their class to works "used … to train" Google's models and are barred from broadening it.

In any event, Plaintiffs' demand is untimely and unjustifiable. Given the massive burdens of training data discovery, and whether such burdens could be justified in a case without a certified class, the parties agreed in May to a framework to reduce burdens: Plaintiffs would select a limited number of exemplar datasets based on data cards and other documents showing the datasets actually used to train, identifying the specific documents describing the selected datasets and linking them to specific at-issue models. That process in May, which the Court considered and approved (*see* ECF No. 155), led to Google making available 19 large datasets (over 1.5 PB) in June. That massive volume of information and myriad datasets amply allowed Plaintiffs to devise

whatever arguments they imagine could support class certification.

The deadline for that certification inquiry has come and gone twice, with Plaintiffs securing modest extensions. Now on the eve of their ***third*** deadline, Plaintiffs demanded and are moving to compel entirely new datasets. Google asked them to provide the same information they had previously—namely, the model they believe each dataset was used to train and the documents describing the dataset and showing that link. Plaintiffs refused to do so, electing instead to forge ahead without even acknowledging the governing process.[6]

Plaintiffs pretend that they only just learned of these datasets. Indeed, they complained in earlier drafts of this motion that "[n]one of the datasets" they now seek "were included on the exemplar data cards Google produced." All but one of the datasets that Plaintiffs now demand (FineWeb, FineWeb2, FineWeb-edu, Common Crawl, The Pile, RedPajama, RedPajama-2), *supra* at 1, do not appear on data cards for a straightforward reason: they were not used to train the models at issue. Regardless, most of these datasets are discussed in another document that Plaintiffs have focused on in depositions. That document, produced to Plaintiffs back on ***May 9***, compares these datasets to others that Google actually used for training, proposes using these datasets in the future (a suggestion that was never approved), and ***expressly disclaims*** actual use of the datasets for training. Plaintiffs could have used this document to request these datasets back in May when they made their other requests but did not, presumably because the document on its face shows that the datasets were ***not*** used to train any models at issue. Had Plaintiffs requested them at the time, Google would have highlighted their non-use. Indeed, Google gave that response for several other datasets Plaintiffs requested in May, and Plaintiffs never disputed the point or further pursued them.

The only other dataset Plaintiffs request [redacted] was one of many used to train Gemini v.2 and v.3, as disclosed in data cards Google produced on April 18 and June 24. Numerous

[6] Plaintiffs did not meet and confer in good faith. They first mentioned these datasets on September 15; by September 24 Plaintiffs were demanding production in a week, without providing any of the information required by the protocol, and sent their portion of the joint letter brief one day later. That rush to filing violates this Court's directives. *See* Standing Order re: Discovery § 7(b); ECF No. 178 at 39:20-40:18.

other internal documents referencing [redacted] were produced as early as April 11. Plaintiffs say they learned of [redacted] only through "later produced evidence," *supra* at 1, but they identify no such evidence—and the record refutes Plaintiffs' lily gilding. Plaintiffs learned of information referencing and discussing [redacted] months ago and simply chose to select many other datasets rather than this one until days before their twice-extended class-certification deadline. That was their choice—not the result of any supposed "tardy custodial productions." *Supra* at 2.[7]

Most importantly, as with much of Plaintiffs' misdirected discovery campaign, they make no showing why this late-breaking demand is relevant and proportional in light of the massive discovery already produced and the substantial additional expense that would be required. They never explain what class-certification argument they hope to advance that they cannot advance without yet another dataset. At this stage, that—not hyperbole—should be the touchstone for additional discovery under Rule 26(b)(1). On that front as well, Plaintiffs' demand fails.

Finally, Plaintiffs' demand is grossly disproportionate. Their assertion that "[p]roduction of the datasets is not burdensome," *supra* at 5, is incorrect. Producing numerous entire datasets, most of which were never used for training, would impose substantial burdens for no meaningful probative value. Preparing the 19 datasets that Google already provided required tracking down massive datasets, converting them into reviewable formats, securely loading the data, and enabling ongoing hosting and support. That process has consumed many hundreds of engineering hours and resulted in substantial, continuing storage and compute costs to facilitate Plaintiffs' access and process resource-intensive queries. Plaintiffs' demand here similarly would likely require weeks of work just to locate and/or restore, gather, prepare, and load the data for review, assuming it can be found. Rule 26(b)(1) requires discovery to be relevant ***and*** proportional; this demand is neither. The Court should deny it and caution Plaintiffs against further make-work motions.

[7] To the extent Plaintiffs are allowed to pursue a request for [redacted] training data, they should be required to specify the particular dataset used for one of the at-issue models, *e.g.*, the [redacted] dataset used to train a Gemini v.2 model or a Gemini v.3 model. An amorphous request for [redacted] data is unproductive and, to the extent it seeks data that was not actually used to train an at-issue model, inconsistent with the complaint and Judge Lee's orders.

Respectfully submitted,

Dated: October 3, 2025

By: */s/ Joseph R. Saveri*

Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Evan A. Creutz (SBN 349728)
Elissa A. Buchanan (SBN 249996)
Aaron Cera (SBN 351163)
Louis Kessler (SBN 243703)
Alex Zeng (SBN 360220)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1505
San Francisco, CA 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
ecreutz@saverilawfirm.com
eabuchanan@saverilawfirm.com
acera@saverilawfirm.com
lkessler@saverilawfirm.com
azeng@saverilawfirm.com

Lesley E. Weaver (SBN 191305)
Anne K. Davis (SBN 267909)
Joshua D. Samra (SBN 313050)
BLEICHMAR FONTI & AULD LLP
1330 Broadway, Suite 630
Oakland, CA 94612
Telephone: (415) 445-4003
lweaver@bfalaw.com
adavis@bfalaw.com
jsamra@bfalaw.com

Gregory S. Mullens (admitted *pro hac vice*)
BLEICHMAR FONTI & AULD LLP
75 Virginia Road, 2nd Floor
White Plains, NY 10603
Telephone: (415) 445-4006
gmullens@bfalaw.com

Ryan J. Clarkson (SBN 257074)
Yana Hart (SBN 306499)
Mark I. Richards (SBN 321252)
CLARKSON LAW FIRM, P.C.
22525 Pacific Coast Highway
Malibu, CA 90265
Telephone: 213-788-4050
rclarkson@clarksonlawfirm.com
yhart@clarksonlawfirm.com
mrichards@clarksonlawfirm.com

Tracey Cowan (SBN 250053)
CLARKSON LAW FIRM, P.C.
95 Third Street, Second Floor
San Francisco, CA 94103
Telephone: (213) 788-4050
tcowan@clarksonlawfirm.com

Brian D. Clark (admitted *pro hac vice*)
Laura M. Matson (admitted *pro hac vice*)
Arielle S. Wagner (admitted *pro hac vice*)
Consuela Abotsi-Kowu (admitted *pro hac vice*)
LOCKRIDGE GRINDAL NAUEN PLLP
100 Washington Avenue South, Suite 2200
Minneapolis, MN 55401
Telephone: (612) 339-6900
Facsimile: (612) 339-0981
bdclark@locklaw.com
lmmatson@locklaw.com
aswagner@locklaw.com
cmabotsi-kowo@locklaw.com

Stephen J. Teti (admitted *pro hac vice*)
LOCKRIDGE GRINDAL NAUEN PLLP
265 Franklin Street, Suite 1702
Boston, MA 02110
Telephone: (617) 456-7701
sjteti@locklaw.com

Counsel for Individual and Representative Plaintiffs and the Proposed Class

Dated: October 3, 2025

Respectfully submitted,

By: */s/ Paul J. Sampson*

Paul J. Sampson
WILSON SONSINI GOODRICH & ROSATI, P.C.
95 S State Street, Suite 1000
Salt Lake City, UT 84111
Telephone: (801) 401-8510
Email: psampson@wsgr.com

David H. Kramer
Maura L. Rees
Qifan Huan
Kelly M. Knoll
WILSON SONSINI GOODRICH & ROSATI, P.C.
650 Page Mill Road
Palo Alto, CA 94304-1050
Telephone: (650) 493-9300
Fax: (650) 565-5100
Email: dkramer@wsgr.com
Email: mrees@wsgr.com
Email: qhuan@wsgr.com
Email: kknoll@wsgr.com

Eric P. Tuttle
Madison Welsh
WILSON SONSINI GOODRICH & ROSATI, P.C.
701 Fifth Avenue, Suite 5100
Seattle, WA 98104-7036
Telephone: (206) 883-2500
Fax: (206) 883-2699
Email: eric.tuttle@wsgr.com
Email: mjwelsh@wsgr.com

Jeremy Paul Auster
WILSON SONSINI GOODRICH ROSATI, P.C.
1301 6th Avenue
New York, NY 10019
212-453-2862
Email: jauster@wsgr.com

Counsel for Defendnats Google LLC And Alphabet Inc.

SIGNATURE ATTESTATION

I, Joseph R. Saveri, am the ECF User whose ID and password are being used to file this document. In compliance with N.D. Cal. Civil L.R. 5-1(i)(3), I hereby attest that the concurrence in the filing of this document has been obtained from the other signatory.

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