

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

IN RE GOOGLE ASSISTANT PRIVACY
LITIGATION.

Case No. 19-cv-04286-BLF (SVK)

**ORDER RE MODIFIED SAMPLING
PROTOCOL**

Pursuant to Judge Freeman’s Order Regarding Motion for Relief from Non-Dispositive Pretrial Order (Dkt. No. 370), the Court has revisited the scope of the October 20, 2022 Sampling Order (Dkt. No. 331) in light of the Class Certification Order (Dkt. 360). Given that only the Purchaser Class was certified to pursue breach of contract and Unfair Competition Law claims predicated on breach of contract and a violation of California Business and Professions Code § 22576, the proportionality analysis has shifted dramatically since the Court issued the initial Sampling Order.

As a result of these developments, the undersigned requested updated briefing regarding an appropriate sampling protocol and held hearings on August 27 and October 3, 2024. *See* Dkt. 438, 443, 451, 452. Through this process, certain parameters on sampling were revised, however, the Parties remained in vastly different universes as to sample size of queries, with Plaintiffs arguing for a sample in excess of 20 million queries (*see, e.g.*, Dkt. 452 at 1) and Google offering to sample merely 12,500 queries. Dkt. 450 at 16:2-9. At the October 3 hearing, the Court admonished both sides that substantial compromise would be required and, if after one final effort the parties could not agree, the Court would determine the sample size. *Id.* at 44:15-19. Both sides were amenable to a final effort and, if they could not agree, accepting a number set by the Court. *Id.* at 43:15-44:3. The Parties’ further submissions reflected some, but not enough,

1 movement, with Plaintiffs reluctantly suggesting 2.6 million (Dkt. 452 at 2) and Google moving to
2 104,000 (Dkt. 451 at 1). Having reviewed the transcripts of the hearings on this issue, the Parties’
3 final submissions, and the declaration of Google’s expert Jonathan Borck (Dkt. 451-2), the Court
4 determines that **a sample size of 260,000** queries is appropriate and proportional to the needs of
5 this case.

6 Accordingly, the Court **ORDERS** as follows:

7 1. Google will produce a sample of speech log data within **45 days** of the date of this
8 Order.

9 2. Over the **26 quarters** that the Parties have agreed are relevant, Google will perform
10 sampling for **1 randomly selected day** per quarter.

11 3. For each sample day, Google will **randomly select 10,000 Google Assistant**
12 **queries** that were (1) hotword-initiated; and (2) made on Google Assistant Enabled Devices
13 manufactured by Google. The total number of queries to be sampled across all 26 quarters is
14 260,000.

15 4. For the set of randomly sampled queries for each sample day, Google will provide
16 (1) the number of queries that did not contain a hot word in the “top hypothesis” field, and (2) the
17 number of those queries that were sent to a human reviewer.

18 5. Google shall also provide Plaintiffs with the raw speech log data for all 104,000
19 queries, with the following modifications:

- 20 a. Google may anonymize the raw data before producing it to Plaintiffs.
- 21 b. Google may redact the content of each query, leaving only the hotword
22 visible. Where there is no hotword, the entire query will be redacted.

23 **SO ORDERED.**

24 Dated: October 17, 2024

25
26
27
28


SUSAN VAN KEULEN
United States Magistrate Judge