

EXHIBIT B

P5RD0pe6B

1 UNITED STATES DISTRICT COURT
2 SOUTHERN DISTRICT OF NEW YORK
-----x

3 THE NEW YORK TIMES COMPANY,

4 Plaintiff,

5 v.

23 CV 11195

6 MICROSOFT CORPORATION, ET AL.,

7 Defendants.

Conference

8
9 -----x

10 Before:

11 HON. ONA T. WANG,
12 Magistrate Judge

13 APPEARANCES

14 SUSMAN GODFREY LLP
Attorneys for The New York Times
15 BY: ALEXANDER P. FRAWLEY
ADNAN MUTTALIB
16 ALEJANDRA C. SALINAS
DAVIDA BROOK
17 ROHIT NATH
JUSTIN NELSON

18 ROTHWELL FIGG ERST & MANBECK, P.C.
Attorneys for New York Times and Daily News
19 BY: JENNIFER MAISEL
20 STEVEN LIEBERMAN
KRISTEN LOGAN

21
22 LOEVY & LOEVY
Attorneys for Center for Investigative Reporting
23 BY: MATTHEW TOPIC

24 BOIES, SCHILLER FLEXNER, LLP
Attorneys for N.D. Cal Plaintiffs
25 BY: JOSHUA M. STEIN
EVAN EZRAY

P5RD0pe6B

1 APPEARANCES, Continued:

2 KEKER, VAN NEST & PETERS
Attorneys for OpenAI
3 BY: KATIE LYNN JOYCE
MICHELLE YBARRA
4 CHRISTOPHER SUN
SARAH SALOMON
5 BILAL MALIK
EDWARD BAYLEY
6 OLIVIA MALONE

7 MORRISON & FOERSTER LLP
Attorneys for OpenAI
8 BY: JOSEPH GRATZ
CAROLYN HOMER
9 ROSE LEE

10
11 LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
Attorneys for Authors Guild and Alter Class Plaintiffs
12 BY: RACHEL GEMAN

13
14 LATHAM & WATKINS LLP
Attorneys for Microsoft
15 BY: ELANA NIGHTINGALE DAWSON
HERMAN YUE

16 FAEGRE DRINKER BIDDLE & REATH LLP
17 Attorneys for Microsoft
BY: JARED BRIANT

18 ORRICK HERRINGTON & SUTCLIFFE LLP
19 Attorneys for Microsoft Corporation
BY: ANNETTE L. HURST
20 SHERYL GARKO

P5RBOEP7

1 THE COURT: Then who retains the data?

2 MR. BAYLEY: They do. It's stored on a production
3 system that OpenAI makes available to then, but the actual data
4 is not something that we'd be able to actually access and
5 retain without building a new system. It is not something that
6 we actually retain ourselves. And, again, that's under our
7 agreements, eyes-off data. The same type of eyes-off data that
8 Microsoft is also not retaining. So this is another area where
9 I'm telling your Honor where we are being treated differently.

10 THE COURT: Microsoft isn't here. You're going to
11 point to them, and I can't ask the questions now. So but just
12 to give me a sense. Microsoft keeps talking about 15
13 petabytes. How does that compare to what has been retained in
14 the consumer gumball machine?

15 MR. BAYLEY: Your Honor, it's difficult because our
16 data, unlike Microsoft's, is kept in a compressed format. It's
17 1.6 petabytes, but that's in a very compressed format. This is
18 a very rough estimate, but my understanding is the compression
19 ratio for this type of data is often 15 to 17 percent. So
20 assuming that's right, the actual uncompressed amount of
21 petabytes would actually be even more than what Microsoft --
22 perhaps like double what Microsoft -- it's a huge, huge amount
23 of data. Like I said, tens of billions of conversations.

24 That is what we're saying when I'm saying, the idea of
25 there being a statistically significant sample. We have a lot

P5RBOEP7

1 to work with here. And if you look at how other cases have
2 addressed this issue, in particular, I think the most pertinent
3 and relevant to this case is one that has just come to ground
4 recently in the *Concord Music Group v. Anthropic* case.

5 *Anthropic* is another LLM company that provides its own AI
6 system Chatbot. They've addressed this very issue about how to
7 sample conversation data in order to figure out what is
8 reasonable and proportionate for the purposes of the case.

9 In that case -- and subject to an order that just came
10 out on Friday, the court has determined that there needs to be
11 a reasonable -- that what is proportionate to the case is a
12 statistically significant sample. She had the experts go,
13 figure out what that means. They came back with competing
14 proposals, and she adopted a proposal somewhere in between the
15 two. And the end result was that a statistically significant
16 sample was five million chats.

17 THE COURT: Five million sorry?

18 MR. BAYLEY: Five million conversations, five million
19 chats. That is -- I can't even do the math, but that is a very,
20 very, very tiny percentage from where we're starting with the
21 tens of billions. And so when I'm saying there's lots of
22 logical inferential loops you have to make here, I'm saying
23 given these vast amount of data that we're starting with, the
24 idea that there's somehow a bias in the sample given how big
25 the sample is, is a little bit difficult for us to come to. We