

Exhibit 5

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 204

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, ET AL.,)
INDIVIDUAL AND REPRESENTATIVE)
PLAINTIFFS,) LEAD CASE NO.
v.) 3:23-cv-03417-VC
META PLATFORMS, INC.,)
DEFENDANT.)

* * * H I G H L Y C O N F I D E N T I A L * * *
* * * A T T O R N E Y S ' E Y E S O N L Y * * *

VIDEO-RECORDED 30 (B) (6) DEPOSITION OF
MICHAEL CLARK
VOLUME III
WEDNESDAY, NOVEMBER 20, 2024
SAN FRANCISCO, CALIFORNIA
9:38 A.M. PST

REPORTED BY AUDRA E. CRAMER, CSR NO. 9901

DIGITAL EVIDENCE GROUP
1730 M Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 314

1 MS. HARTNETT: Objection to the form.

2 THE WITNESS: The removal of that data
3 kept there from being additional noise
4 introduced from repetitive phrases and from
5 things that would introduce other noise or
6 problems into the model itself.

7 BY MS. POUEMYMIROU:

8 Q. And would you say that removal of that
9 facilitates training?

10 A. Removal of that data is one step in the
11 data cleanup phase, which is one of many pieces
12 to facilitate training.

13 Q. Okay. So in the opening of this
14 document, you have Sergey Edunov and Melanie
15 Kambadur noted.

16 And what I was hoping you could explain
17 is what is meant by this specific language: "We
18 need to make sure source metadata does not cause
19 problems in training, stability, performance or
20 leakage of sensitive info. We have experimental
21 data now showing basically no negative effects
22 for the v1 unhashed source metadata, but since

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 315

1 this is small-scale, noisy evals, it's
2 unfortunately not 100 percent conclusive.
3 Intuitively, adding 5 to 10 percent to token
4 size could be a risk."

5 And this is the sentence in particular
6 I'm interested in: "Also, unhashed source
7 metadata is not acceptable from an IP-brand
8 safety leakage perspective IMO. We have to be
9 basically 100 percent confident no one can
10 extract this data, and based on previous
11 memorization work, I do not think we can be."

12 So can you walk me through what they're
13 talking about.

14 MS. HARTNETT: Object to the form.

15 THE WITNESS: I was trying to find the
16 exact line.

17 Was that the part from Melanie, or was
18 that the --

19 BY MS. POUYMIROU:

20 Q. It looks like -- I mean, it's
21 "Melanie:" and then @sergeyedunov?

22 MS. HARTNETT: I think she's reading on

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 316

1 the flip side, which is the --

2 THE WITNESS: Oh, sorry.

3 MS. HARTNETT: No, no, no. On the
4 first side, but the sentence that starts on the
5 flip side.

6 THE WITNESS: Thank you. I was lost
7 for a second there.

8 MS. HARTNETT: Yeah, there's an "also"
9 sentence there in the middle there.

10 THE WITNESS: Can I just ask: Do we
11 have the context before this to know which
12 source metadata was being referred to?

13 BY MS. POUYMIROU:

14 Q. No. We just have this.

15 A. Okay.

16 Okay. I asked the question on context
17 around this to know if source metadata referred
18 to dataset or individual elements in datasets or
19 something else, because metadata can be attached
20 to a variety of things.

21 When I -- and so not knowing which one
22 it is referring to specifically, I think the --

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 348

1 sharing the content outside. This could be
2 legally not okay."

3 A. Okay.

4 Q. So maybe I'm misunderstanding, but I
5 thought you said that Bashlykov said that they
6 did use torrents, and I'm just trying to
7 confirm -- this document suggest that they did a
8 direct file upload -- if that's different than
9 what he testified to. I'm just trying to
10 understand.

11 A. Let me read in context the other parts
12 of his notes here.

13 Okay. Apologize. I was confused by
14 the page order.

15 Q. That's okay.

16 A. Do you mind asking that question once
17 again?

18 Q. Yeah. So I believe you've testified
19 that you spoke with Bashlykov and that he said
20 that they downloaded LibGen through torrent.

21 Is that correct?

22 A. That is correct.

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 349

1 Q. This says, "Decided to go with direct
2 file upload without using torrents for the
3 following reasons: Using torrents would entail
4 seeding the files, sharing the content outside.
5 This could be legally not okay."

6 So how did Meta obtain and copy LibGen?

7 MS. HARTNETT: Object to the form.

8 THE WITNESS: So I spoke with
9 Bashlykov. We used libtorrent and the torrent
10 protocol for downloading. We had tried to
11 download a few different ways. One of them
12 included doing it via direct downloads where
13 there were individually hosted files in
14 different places and saw many issues with that
15 where files wouldn't exist, it was too slow or
16 would error out.

17 Also tried -- well, actually, the other
18 only other option was downloading via torrent,
19 and --

20 BY MS. POUYMIROU:

21 Q. And that's what you did?

22 A. That's what we did. And the library

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 350

1 that we used using, LibGen -- sorry --
2 libtorrent for downloading LibGen, Bashlykov
3 modified the config setting so that the smallest
4 amount of seeding possible could occur.

5 Q. What does that mean?

6 A. When you use a torrent protocol, part
7 of the configuration on how torrents work is you
8 can only download as long as you offer to
9 participate in the torrent network in some way,
10 and seeding means that you have to open up some
11 amount of sharing of the torrent file while
12 you're downloading.

13 Q. So why did Meta decide to minimize
14 seeding the files?

15 Is it because it was concerned that
16 that was legally not okay, to quote this
17 document?

18 MS. HARTNETT: Objection to the form.

19 And to the extent it calls for any
20 attorney-client privilege, please don't reveal
21 that.

22 THE WITNESS: Yeah, I do not know what

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 351

1 Nikolay means in here by "legally not okay." I
2 know that legal did provide advice, which I
3 can't share. I can share what the setting was.

4 BY MS. POUEMYIROU:

5 Q. Is Bashlykov a lawyer?

6 A. No.

7 Q. Okay. So you ended up using torrent,
8 but you minimized seeding the files?

9 A. To the lowest possible setting, yes.

10 Q. Are you familiar with anti-leeching
11 protocol?

12 A. I've heard the term before.

13 In which context?

14 Q. In the context of -- with respect to
15 seeding the files.

16 Is there a relationship what Meta was
17 doing by minimizing the seeding of file with
18 leeching?

19 A. Could you be more specific?

20 Q. How do you understand anti-leeching?

21 A. I don't -- I didn't prepare on it
22 today. I would only know from my just own