

1
2
3
4
5
6
7 UNITED STATES DISTRICT COURT
8 WESTERN DISTRICT OF WASHINGTON
9 AT SEATTLE

10 STEVEN VANCE, et al.,

11 Plaintiffs,

12 v.

13 MICROSOFT CORPORATION,

14 Defendant.

CASE NO. C20-1082JLR

ORDER ON MICROSOFT'S
MOTION FOR SUMMARY
JUDGMENT

15 **I. INTRODUCTION**

16 Before the court is Defendant Microsoft Corporation's ("Microsoft") renewed
17 motion for summary judgment. (Mot. (Dkt. # 127); Reply (Dkt. # 138).) Plaintiffs
18 Steven Vance and Tim Janecyk (collectively, "Plaintiffs") oppose Microsoft's motion.
19 (Resp. (Dkt. # 135¹)). The court has considered the motion, all materials submitted in

20
21 ¹ Plaintiffs originally filed their response under seal because it relied on and cited
22 documents that Microsoft had marked confidential; they also filed a redacted version of their
response. (Mot. to Seal (Dkt. # 134); Redacted Resp. (Dkt. # 132).) Because Microsoft did not
oppose unsealing the response and the documents, the court denied Plaintiffs' motion to seal and

1 support of and in opposition to the motion, and the governing law. Being fully advised,²
2 the court GRANTS Microsoft's motion for summary judgment.

3 II. BACKGROUND

4 The court sets forth the factual and procedural background of this case below.

5 A. Factual Background

6 1. The Diversity in Faces ("DIF") Dataset

7 Plaintiffs are longtime Illinois residents who, beginning in 2008, uploaded digital
8 photographs, including photos of themselves, to Flickr, a photo-sharing website. (*See*
9 Compl. (Dkt. # 1) ¶¶ 6-7, 28, 66-67, 75; Vance Dep.³ at 9:15-10:9; Janecyk Dep.⁴ at
10 39:7-40:1.) In 2014, Yahoo!, Flickr's then-parent company, publicly released a dataset of
11 about 100 million photographs that had been uploaded to Flickr's website between 2004
12

13 directed the clerk to remove the seal on Plaintiffs' responsive brief and the confidential
14 documents. (Mot. to Seal Resp. (Dkt. # 136); 7/11/22 Order (Dkt. # 137).) Accordingly, the
15 court cites the unredacted version of Plaintiffs' response in this order.

16 ² Both parties request oral argument on the motion (*see* Mot. at 1; Resp. at 1). The court,
17 however, concludes that oral argument would not be helpful to its disposition of the motion. *See*
18 Local Rules W.D. Wash. LCR 7(b)(4).

19 ³ Both parties have submitted excerpts from Mr. Vance's deposition. (*See* Berger Decl.
20 (Dkt. # 86) ¶ 2, Ex. 1; 7/1/22 Lange Decl. (Dkt. # 132-1) ¶ 2, Ex. 1.) For ease of reference, the
21 court cites directly to the page and line number of the deposition.

22 The court notes that Plaintiffs did not highlight the portions of the deposition transcripts
that they referred to in their pleadings as required by Local Civil Rule 10(e)(10). *See* Local
Rules W.D. Wash. LCR 10(e)(10) ("All exhibits [submitted in support of or in opposition to a
motion] must be marked to designate testimony or evidence referred to in the parties' filings.")
The court directs Plaintiffs' counsel to review the local rules regarding marking exhibits before
making any further filings.

⁴ Both parties have submitted excerpts from Mr. Janecyk's deposition. (*See* Berger Decl.
¶ 3, Ex. 2; 7/1/22 Lange Decl. ¶ 3, Ex. 2.) For ease of reference, the court cites directly to the
page and line number of the deposition.

1 and 2014 (the “YFCC-100M Dataset”). (See Merler Decl. (Dkt. # 85) ¶ 3, Ex. A
2 (“*Diversity in Faces*”) at 2.) The YFCC-100M Dataset included photos uploaded by both
3 Plaintiffs. (See Vance Dep. at 179:22-23; Janecyk Dep. at 95:22-24.)

4 Before 2018, “there was an industry-wide problem with many facial recognition
5 systems’ ability to accurately characterize individuals who were not male and did not
6 have light colored skin tones.” (Merler Decl. ¶ 4.) As a result, “the facial recognition
7 systems and algorithms associated with those facial recognition systems were trained in
8 such a way that the systems were able to accurately characterize a white, light skinned
9 male subject, but the technology suffered from inaccuracies when it had to characterize a
10 non-male or a person with darker skin tones.” (*Id.*) Seeking to “advance the study of
11 fairness and accuracy in face recognition technology,” researchers working for
12 International Business Machines Corporation (“IBM”)⁵ used one million of the photos in
13 the YFCC-100M Dataset to develop the Diversity in Faces (“DiF”) Dataset at issue in
14 this case. (*Id.* ¶ 5; *Diversity in Faces* at 2, 7.) The researchers implemented ten “facial
15 coding schemes” to measure aspects of the facial features of the individuals pictured in
16 the photos, such as “craniofacial distances, areas and ratios, facial symmetry and contrast,
17 skin color, age and gender predictions, subjective annotations, and pose and resolution.”
18 (*Diversity in Faces* at 9.) A statistical analysis of these coding schemes “provided insight
19 into how various dimensions . . . provide indications of dataset diversity.” (Merler
20

21 ⁵ All of the researchers involved in creating the DiF Dataset were based in and worked
22 out of IBM’s office in Yorktown Heights, New York; and the work was performed on and stored
on IBM Research computer servers in Poughkeepsie, New York. (*Id.* ¶ 8.) None of the work
involved computers or systems located in Illinois. (*Id.*)

1 Decl. ¶ 6.) The coding schemes implemented by the IBM researchers were intended to
2 enable other researchers to develop techniques to estimate diversity in their own datasets,
3 with the goal of mitigating dataset bias, and were “never intended to identify any
4 particular individual.” (*Id.* ¶ 7.) Rather, the coding schemes were “purely descriptive
5 and designed to provide a mechanism to evaluate diversity in the dataset.” (*Id.*)

6 IBM provided the DiF Dataset free of charge to researchers who filled out a
7 questionnaire and submitted it to IBM via email. (*Id.* ¶¶ 4, 9.) The questionnaire
8 required the researcher to verify

9 (i) that he/she would only use the DiF Dataset for research purposes, and
10 (ii) that he/she had read and agreed to the DiF Dataset terms of use, which
11 made clear that the DiF Dataset could only be used for non-commercial,
research purposes and prohibited using the DiF Dataset to identify any
individuals in images associated with URLs in the DiF Dataset.

12 (*Id.* ¶ 9; *see also id.* ¶ 11, Ex. H (DiF Dataset terms of use).) After verifying that a
13 request was for a “legitimate research purpose,” IBM researcher Dr. Michele Merler sent
14 the DiF Dataset to the requesting researcher “via an email that included a link to a
15 temporary Box folder that contained the DiF Dataset.” (Merler Decl. ¶ 10.)

16 2. Plaintiffs’ Photos in the DiF Dataset

17 The DiF Dataset includes at least 61 of the nearly 19,000 public photos that Mr.
18 Vance uploaded to Flickr. (Vance Dep. at 179:22-23, 210:19-24.) Mr. Vance appears in
19 some of the photos in the DiF Dataset; other photos depict people whose state of
20 residence was unknown to Mr. Vance and at least one depicts individuals who themselves
21 were unknown to Mr. Vance. (*Id.* at 132:4-14; 154:5-16.)

1 The DiF Dataset includes 24 of the 1,669 public photos that Mr. Janecyk uploaded
2 to Flickr. (Janecyk Dep. at 74:21-24, 95:22-96:1.) Mr. Janecyk appears in at least one of
3 the photos. (*Id.* at 99:21-100:6.) Because Mr. Janecyk photographed people on the
4 streets of Chicago, however, he does not know the names or places of residence of the
5 individuals depicted in most of his photos. (*Id.* at 45:16-46:19, 98:8-100:13,
6 167:11-168:15, 228:19-21.)

7 3. Microsoft's Downloads of the DiF Dataset

8 Two individuals affiliated with Microsoft downloaded the DiF Dataset in February
9 2019: contractor Benjamin Skrainka and Microsoft Research intern Samira Samadi.
10 (Skrainka Decl. (Dkt. # 87) ¶ 5; Samadi Decl. (Dkt. # 88) ¶¶ 5-6.) The court describes
11 their interactions with the DiF Dataset below.

12 *a. Benjamin Skrainka*

13 Between September 7, 2018, and August 1, 2019, Mr. Skrainka worked as an
14 independent contractor for Neal Analytics, LLC, a Washington-based consulting firm,
15 through which he contracted as a vendor to Microsoft. (Skrainka Decl. ¶ 2; Skrainka
16 Dep.⁶ at 91:7-24, 111:8-23.) During this period, Mr. Skrainka provided support for a
17 project, Azure Intelligent Storage (“AIS”), for Microsoft. (Skrainka Decl. ¶ 3.) His work
18 related to defining a benchmark protocol for evaluating a third-party facial recognition
19 technology that Microsoft was considering acquiring. (*Id.*; Kasap Decl. (Dkt. # 91)

20
21 ⁶ Both parties have submitted excerpts from Benjamin Skrainka's deposition. (*See*
22 5/19/22 Wiese Decl. (Dkt. # 129) ¶ 2, Ex. 1; 7/1/22 Lange Decl. ¶ 12, Ex. 11; 7/29/22 Wiese
Decl. (Dkt. # 139) ¶ 2, Ex. 9.) For ease of reference, the court cites directly to the page and line
number of the deposition.

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.