

No. 24-145

**IN THE UNITED STATES COURT OF APPEALS
FOR THE NINTH CIRCUIT**

MICHELE RONK,

Plaintiffs-Appellant,

v.

KATHRYN ELIZABETH HUDSON, ET AL.,

Defendants-Appellees.

On Appeal from the United States District Court
for the Central District of California
No. 2:20-cv-09843
Hon. Fernando L. Aenlle-Rocha

APPELLANTS' OPENING BRIEF

Steven T. Lowe, Esq.
Lowe & Associates, P.C.
8383 Wilshire Blvd. #1038
(310) 477-5811
Email: steven@lowelaw.com

Attorney for Appellant
Michele Ronk

TABLE OF CONTENTS

| | Page(s) |
|--|----------------|
| TABLE OF AUTHORITIES | iii-viii |
| I. INTRODUCTION | 1 |
| II. JURISDICTIONAL STATEMENT | 5 |
| III. ISSUES PRESENTED | 5 |
| IV. STATEMENT OF THE CASE..... | 6 |
| A. PROCEDURAL HISTORY | 6 |
| B. STATEMENT OF FACTS..... | 7 |
| 1. THE PARTIES' WORKS..... | 7 |
| 2. FACEBOOK SHARED ITS USERS' DATA WITH PARTNERS | 8 |
| a. FACEBOOK'S DEMONSTRATED DATA SHARING ACTIVITY | 12 |
| 3. FACEBOOK AND UMG ARE STRATEGIC PARTNERS SINCE 2017..... | 15 |
| 4. AVAILABILITY OF ARTIFICIAL INTELLIGENCE PROGRAMS..... | 17 |
| a. GPT-1 | 18 |
| b. GPT-2 | 20 |
| c. Bidirectional Encoder Representations from Transformers (BERT) | 21 |
| d. Text-to-Text Transfer Transformer (T5) | 22 |

| | |
|--|----|
| e. XLNet | 24 |
| f. Robustly Optimized BERT Pretraining Approach (RoBERTa) | 26 |
| g. DistilBERT | 27 |
| h. ALBERT..... | 28 |
| i. Stanford CoreNLP | 30 |
| j. Availability of Artificial Intelligence Programs Prior to July 2020 | 31 |
| 5. SUMMARY OF FACTS AND ALLEGED CHAIN OF EVENTS BETWEEN APPELLANT’S WORK AND APPELEES’ ACCESS. | 34 |
| V. SUMMARY OF THE ARGUMENT | 35 |
| VI. ARGUMENT | 36 |
| A. STANDARD OF REVIEW | 36 |
| B. GIVEN THE CURRENT TECHNOLOGICAL LANDSCAPE, THE COURT SHOULD EXPAND THE CIRCUMSTANCES WHERE THERE IS A ‘REASONABLE POSSIBILITY OF ACCESS’ | 37 |
| C. THE COURT SHOULD RECONSIDER WHAT ‘REASONABLE POSSIBILITY OF ACCESS’ MEANS IN LIGHT OF MODERN DATA SHARING AND THE AVAILABILITY OF ARTIFICIAL INTELLIGENCE PROGRAMS | 40 |
| D. THE USE OF ARTIFICIAL INTELLIGENCE PROGRAMS HAS MASSIVE IMPLICATIONS FOR COPYRIGHT INFRINGEMENT CASES | 45 |
| E. THE CHAIN OF EVENTS AS DESCRIBED HEREINABOVE SHOWS A REASONABLE POSSIBILITY OF ACCESS..... | 47 |
| F. THE DISTRICT COURT ABUSED ITS DISCRETION IN FAILING | |

| | |
|---|----|
| TO GRANT APPELLANTS LEAVE TO AMEND THE COMPLAINT..... | 49 |
| VII. CONCLUSION..... | 50 |
| STATEMENT OF RELATED CASES | |
| CERTIFICATE OF COMPLIANCE | |
| CERTIFICATE OF SERVICE | |

TABLE OF AUTHORITIES

| Cases | Page(s) |
|---|----------|
| <i>Allen v. Ghoulish Gallery</i> , No. 06-cv-371 NLS, 2007 U.S. Dist. LEXIS 86224 (S.D. Cal. Nov. 20, 2007) | 41 |
| <i>Amerisource Bergen Corp. v. Dialysist W., Inc.</i> , 465 F.3d 946, 949, 957 (9th Cir. 2006)..... | 37 |
| <i>Arnstein v. Porter</i> , 154 F.2d 464 (2d Cir. 1946) | 41 |
| <i>Astor-White v. Strong</i> , 733 F. App'x 407 (9th Cir. 2018) | 49-50 |
| <i>Baccei v. United States</i> , 632 F.3d 1140 (9th Cir. 2011) | 39 |
| <i>Bell Atlantic Corp. v. Twombly</i> , 550 U.S. 550 (2007)..... | 36 |
| <i>Campbell v. Facebook, Inc.</i> 951 F.3d 1106 (9th Cir. 2020) | 8, 13-14 |
| <i>Chabon v. Meta Platforms, Inc.</i> , 3:23-cv-04663 (N.D. Cal. 2023)..... | 47, 48 |
| <i>Conf. Archives, Inc. v. Sound Images, Inc.</i> , No. 3:2006-76, 2010 U.S. Dist. LEXIS 46955 (W.D. Pa. Mar. 31, 2010) | 42 |
| <i>DuMond v. Reilly</i> , No. CV 19-8922-GW-AGR _x , 2021 U.S. Dist. LEXIS 199398, (C.D. Cal. Jan. 8, 2021) | 41 |
| <i>Gray v. Perry</i> , No. 2:15-CV-05642-CAS-JC _x , 2020 U.S. Dist. LEXIS 46313 (C.D. Cal. Mar. 16, 2020) | 41 |

| | |
|---|-------------------|
| <i>Green Bullion Financial Servs., LLC v. Money4Gold Holdings, Inc.</i> , 639 F. Supp. 2d 1356 (S.D. Fla. 2009)..... | 41 |
| <i>In the Matter of Facebook, Inc.</i> , 19-cv-2184 (D.C. 2019)..... | 9, 10, 49 |
| <i>Kadrey v. Meta Platforms, Inc.</i> , 3:23-cv-03417 (N.D. Cal. 2023)..... | 47, 48 |
| <i>Kimes v. Stone</i> , 84 F.3d 1121 (9th Cir. 1996)..... | 39 |
| <i>Leadsinger, Inc. v. BMG Rights Publ'g</i> , 512 F.3d 522 (9th Cir. 2008)..... | 50 |
| <i>Manzarek v. St. Paul Fire & Marine, Ins. Co.</i> , 519 F.3d 1025 (9th Cir. 2008)..... | 50 |
| <i>Narayanan v. British Airways</i> , 747 F.3d 1125 (9th Cir. 2014)..... | 38 |
| <i>Patel v. Facebook, Inc.</i> , 932 F.3d 1264 (9th Cir. 2019)..... | 8 |
| <i>Parks Sch. of Business, Inc. v. Symington</i> , 51 F.3d 1480 (9th Cir. 1995) 51 F.3d 1480 (9th Cir. 1995)..... | 39 |
| <i>Pullman-Standard v. Swint</i> , 456 U.S. 273 (1982)..... | 38 |
| <i>Sec. Litig. v. Facebook Inc.</i> , 87 F.4th 934 (9th Cir. 2017)..... | 9, 11, 13, 16, 17 |
| <i>Shaw v. Lindheim</i> , 919 F.2d 1352 (9 th Cir. 1990)..... | 36-37 |
| <i>Singleton v. Wulff</i> , 428 U.S. 106 (1976)..... | 39 |
| <i>Skidmore v. Led Zeppelin</i> , 905 F.3d 1116 (9th Cir. 2020)..... | 5, 36, 40, 41, 42 |
| <i>TeamLab Inc. v. Museum of Dream Space, LLC</i> , 650 F. Supp. 3d 934 (C.D. Cal. 2023)..... | 41 |

| | |
|---|----------------------------|
| <i>Thaler v. Perlmutter</i> 2023 U.S. Dist. LEXIS 145823 | 2, 45 |
| <i>Zucco Partners, LLC v. Digimarc Corp.</i> , 552 F.3d 981 (9th Cir. 2009) | 36 |
| Statutes | |
| 28 U.S.C. § 1331 | 5 |
| 28 U.S.C. § 1338(a)..... | 5 |
| 28 U.S.C. § 1291 | 5 |
| 28 U.S.C. § 2107(c)..... | 5 |
| Rules | |
| Fed. R. Civ. P. 12(b)(6)..... | 1, 2, 5, 6, 36, 47, 48, 50 |
| Other Authorities | |
| Ashish Vaswani et al., <i>Attention Is All You Need</i> , arXiv, (Jun. 12, 2017), https://arxiv.org/pdf/1706.03762.pdf | 18, 20 |
| Ben Lutkevich, <i>What is Hugging Face?</i> , TECHTARGET, https://www.techtargget.com/whatis/definition/Hugging-Face (last updated Sept. 2023) | 32 |
| Christopher Buccafusco, <i>There’s No Such Thing as Independent Creation, and It’s a Good Thing, Too</i> , 64 Wm. & Mary L. Rev. 1617 (2023), https://wmlawreview.org/sites/default/files/wmlr_64-6_Buccafusco-pgs1617- 1676.pdf | 43, 44 |
| Christopher D. Manning et al., <i>The Stanford CoreNLP Natural Language Processing Toolkit</i> , The Stanford Natural Language Processing Group, (2014), https://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf | 30, 31 |
| Christina R. Dimeo, <i>Rethinking Music Copyright Infringement in the Digital World: Proposing a Streamlined Test After the Demise of the Inverse Ratio Rule</i> , 55 Univ. Rich. L. Rev. 1077, 1102-03 (2021), | |

| | |
|--|---------------|
| https://scholarship.richmond.edu/lawreview/vol55/iss3/9 | 44 |
| Colin Raffel et al., <i>Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer</i> , arXiv, (Oct. 23, 2019), https://arxiv.org/pdf/1910.10683.pdf | 23 |
| Congressional Research Services (September 29, 2023), <i>Generative Artificial Intelligence and Copyright Law</i> (CRS Report No. LSB10922), https://crsreports.congress.gov/product/pdf/LSB/LSB10922 | 45, 46, 47 |
| <i>Datasets</i> , HUGGING FACE, https://huggingface.co/datasets (last visited Apr. 20, 2024) | 32, 33 |
| Emily Ranger-Murdock, “ <i>Blurred Lines</i> ” to “ <i>Stairway to Heaven</i> ”: <i>Applicability of Selection and Arrangement Infringement Actions in Musical Compositions</i> , 67 <i>UCLA L. Rev.</i> 1066 (2020) | 43 |
| Jacob Devlin et al., <i>BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding</i> , arXiv, (Oct. 11, 2018), https://arxiv.org/pdf/1810.04805.pdf | 21, 22 |
| <i>Facebook and Universal Music Group Strike Unprecedented Global Agreement</i> , UMG (Dec. 21, 2017) https://www.universalmusic.com/facebook-universal-music-group-strike-unprecedented-global-agreement/ | 15 |
| <i>Facebook’s Data-Sharing Deals Exposed</i> , BBC (Dec. 19, 2018), https://www.bbc.com/news/technology-46618582 | 12, 16 |
| Federal Trade Commission, <i>Dissenting Statement of Commissioner Rebecca Kelly Slaughter Regarding the Matter of Facebook, Inc.</i> , at 1 (July 24, 2019), https://www.ftc.gov/system/files/documents/public_statements/1536946/092_3184_facebook_majority_statement_7-24-19.pdf | 11, 12, 15-16 |
| Federal Trade Commission, <i>Dissenting Statement of Commissioner Rohit Chopra Regarding the Matter of Facebook, Inc.</i> , at 1 (July 24, 2019), https://www.ftc.gov/system/files/documents/public_statements/1536946/092_3184_facebook_majority_statement_7-24-19.pdf | 11, 12, 13 |
| Federal Trade Commission, <i>Statement of Chairman Joe Simons and Commissioners Noah Joshua Phillips and Christine S. Wilson Regarding the Matter of Facebook, Inc.</i> , at 1 (July 24, 2019), | |

| | |
|---|--|
| https://www.ftc.gov/system/files/documents/public_statements/1536946/092_3184_facebook_majority_statement_7-24-19.pdf | 10, 11 |
| Gabriel J.X. Dance, et al., <i>As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants</i> , The New York Times (Dec. 18, 2018), https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html | 13, 17 |
| <i>Hugging Face – The AI Community Building the Future</i> , HUGGING FACE, https://huggingface.co/ (last visited Apr. 19, 2024) | 32 |
| <i>Models</i> , HUGGING FACE, https://huggingface.co/models (last visited Apr. 20, 2024). | 32 |
| Motley Fool, <i>Facebook Scores Deal With Last Major Record Label</i> , Fox Business (March 9, 2018), https://www.fool.com/investing/2018/03/09/facebook-scores-deal-with-last-major-record-label.aspx | 15 |
| Partha Pratim Ray, <i>ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope</i> , ScienceDirect, (Apr. 14, 2023), https://www.sciencedirect.com/science/article/pii/S266734522300024X#bib152 .. | 18, 19, 20, 21, 22, 23, 24, 25, 26, 31 |
| <i>Pipelines for Inference</i> , HUGGING FACE, https://huggingface.co/docs/transformers/pipeline_tutorial (last visited Apr. 20, 2024) | 33 |
| <i>Posts</i> , HUGGING FACE, https://huggingface.co/posts (last visited Apr. 19, 2024) | 33 |
| Rebecca Alba, <i>Closing the Stairway to Heaven: The Ninth Circuit Disincentivizes Music Creation</i> , 23 Tul. J. Tech. & Intell. Prop. 197 (2021), https://journals.tulane.edu/TIP/article/view/3656 | 43, 44 |
| SaiPranay Vellala, <i>Conquering Copyright: Why Copyright Needs to be Modernized Based on Practical Illustrations of Inconsistent Copyright Precedent</i> , 56 Akron L. Rev. 409 (2022), https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=2572&context=akronlawreview | 44 |

Tony Romm, *Facebook says a new bug allowed apps to access private photos of up to 6.8 million users*, The Washington Post (Dec. 14, 2018) <https://www.washingtonpost.com/technology/2018/12/14/facebook-says-new-bug-allowed-apps-access-private-photos-up-million-users/> 14

Transformers, HUGGING FACE, <https://huggingface.co/docs/transformers/en/index> (last visited Apr. 19, 2024) 32, 33

USPTO, Public Views on Artificial Intelligence and Intellectual Property Policy (October 2020), https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf 16, 46, 47

Victor Sanh et al., *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*, arXiv, (Oct. 2, 2019), <https://arxiv.org/pdf/1910.01108.pdf> 27, 28

Yinhan Liu et al., *RoBERTa: A Robustly Optimized BERT Pretraining Approach*, arXiv, (Jul. 26, 2019), <https://arxiv.org/pdf/1907.11692.pdf> 26

Zhenzhong Lan et al., *ALBERT: A Lite BERT for Self-supervised Learning of Language Representations*, arXiv, (Sep. 26, 2019) 28, 29

Zhilin Yang et al., *XLNet: Generalized Autoregressive Pretraining for Language Understanding*, arXiv, (Jun. 19, 2019), <https://arxiv.org/pdf/1906.08237.pdf> 24, 25

..... 24, 25

Appellate Practice in Federal and State Courts § 3.05 38

4 Federal Litigation Guide § 48.123 (2024) 38

I. INTRODUCTION

Michele Ronk (“Appellant” or “Ronk”) appeals from the District Court’s Rule 12(b)(6) dismissal of her first amended complaint for copyright infringement. Appellant contends that the production and exploitation of the lyrical content of the musical composition embodied on the master sound recording, *Smile* as performed by Defendant Kathryn Elizabeth Hudson p/k/a Katy Perry (“Hudson” or “Perry”), and which writer’s credits are attributed to Defendants Hudson, Brittany Hazzard p/k/a Starrah, Ferras Alqaisi, Oliver Goldstein, p/k/a Olige, Josh Abraham, and Robert Mandell p/k/a G Koop, and released by Defendant Capitol Records, LLC, a subsidiary of Defendant Universal Music Group, Inc. (collectively “Defendants” or “Appellees”) infringed upon Appellant’s original copyrighted work entitled *Upgraded 2.0* (the “Work”).

The issue front and center in the instant Appeal, is what constitutes a “reasonable possibility of access” in a copyright infringement analysis. The jurisprudence on the issue is abundant to say the least; however, there is shockingly little precedent that addresses today’s digital landscape; not just an online, digitally connected landscape, but one where people use bots/AI programs to do the searching, accessing, and generation of derivative content for them. In this brave new world, billions of people upload enormous amounts of information online, including original creative works; the sites to which they post (e.g., Facebook), demonstrably hoard that data and monetize it by selling, licensing, or

making it available to ‘strategic partners.’ In some cases, the sheer volume of that data makes it ripe for mishandling, causing it to be leaked or unknowingly provided to third parties. But what are the chances that even if someone had access to this enormous amount of data, he or she could find or use a single page of information out of it? Is it a proverbial needle in a haystack? Maybe a human couldn’t do it alone, but with the right tools, it’s not only possible, but plausible that AI technology could do so (at the hands of a human). *Thaler v. Perlmutter*, 2023 U.S. Dist. LEXIS 145823, at *20 (granting summary judgment denying copyright registration for AI generated work “created absent any human involvement”).

Modern artificial intelligence programs make it commonplace to digest hopelessly massive data sets, and find, incorporate or otherwise utilize that data in furtherance of a user’s goal, for example, writing lyrics to a song. Further, there is a deluge of evidence that data uploaded to certain social media sites, such as Facebook, is not secure or responsibility managed – thus, if a third party was to have access to that data and utilize one of the many publicly available artificial intelligence programs in order to create a derivative work therefrom, the facts surrounding that access would be so obscure and hidden, that it would prevent a plaintiff from proceeding past the 12(b)(6) stage without the opportunity for discovery. It necessarily follows that in *this* modern world, the reasonability of access should be apparent if alleged that (a) a person shared original creative work

to particular online site; (b) that data was likely to have been compromised by being provided to a third party; (c) that third party had access to and utilized one or more artificial intelligence programs (or other data scraping technologies); and (d) “created” and exploited a derivative work based upon the subject data gathered. A brief overview of the facts giving rise to this case are as follows:

In or about 2014, Appellant Michele Ronk created the unique and creative musical composition (lyrics) for a song she entitled *Upgraded 2.0* (ER_076 ¶ 18) (the “Work”). On September 23, 2014, Appellant shared her Work to her Facebook page via a status update containing an image of her desk and the text of the aforementioned lyrics (ER_077-78). On or about July 10, 2020, Appellee Hudson, a highly successful pop star professionally known as Katy Perry, released the master recording, *Smile* (ER_083 ¶ 31; E.R_086 ¶ 41). Perry is signed to Appellee Capitol Records, LLC, which is owned by Appellee Universal Music Group (ER_083 ¶ 31). On July 11, 2020, Perry posted a picture to her Instagram page of a computer screen showing some of the lyrics from *Smile*; Appellant, who follows Perry on Instagram, immediately recognized prominent similarities between the *Smile* lyrics and that of the Work. (ER_086 ¶ 41). Upon first glance, the issue of how Appellees accessed Appellant’s Work was not readily apparent. However, following research into Facebook’s mishandling and monetization of its users’ data, and Appellees’ likely opportunity to access Plaintiffs’ work therefrom, the chain of events not only came into focus, but revealed a glaring and problematic

loophole in which gatherers, distributors and consumers of big data (e.g., Facebook and UMG) may facilitate copyright infringement with impunity under a veritable shroud of darkness. (ER_087 ¶ 44 – ER_094 ¶ 56).

The Central District Court granted Defendants’ Motions to Dismiss solely on the grounds that the allegations in Appellant’s First Amended Complaint regarding Defendants’ access was too speculative to establish more than a bare possibility that Defendants obtained access to her Work (ER_026; ER_030). While Appellant offered several theories in her First Amended Complaint regarding access (i.e., widespread dissemination and striking similarity), the instant Appeal addresses the Central District Court’s ruling only with respect to its *factual* finding that Plaintiff’s theory regarding the chain of events linking plaintiff’s work and defendant’s access is too speculative (ER_027). Further, although Plaintiff set forth multiple ‘chain of events’ theories in the FAC, the instant Appeal focuses solely on one: (1) Plaintiff posted the Work to Facebook; (2) Facebook improperly shared data, which likely included Plaintiff’s Work, to third parties, which likely included Defendant UMG; (3) Defendant UMG utilized the data by *inter alia* providing it to Capitol records and its artists, writers, producers, etc.; (4) Defendant utilized software program(s) (which may have included artificial intelligence tools) to find and incorporate Plaintiff’s Work into a substantially similar derivative work, that is Perry’s *Smile*.

/ / /

II. JURISDICTIONAL STATEMENT

The district court had jurisdiction of the case that is docketed as No. 2:20-cv-09843 pursuant to 28 U.S.C. § 1331 (“federal question jurisdiction”) and § 1338(a) (actions arising under any Act of Congress relating to copyright claims). The Court of Appeals has jurisdiction of this appeal pursuant to 28 U.S.C. § 1291. The final judgment that is being appealed from all disposed of issues in this cause and was entered on November 13, 2023. The Notice of Appeal was filed on December 12, 2023. The claimed appeal is timely pursuant to 28 U.S.C. § 2107(c). No motion for a new trial or alteration of the judgment or any other motion that would have tolled the time to appeal was filed. This appeal is from a final order that disposed of all parties’ claims.

III. ISSUES PRESENTED

Did the trial court properly grant a Federal Rule of Civil Procedure 12(b)(6) motion to dismiss in favor of Appellees based upon a factual finding as to the lack of access?

- A. Did the District Court err in failing to find a reasonable possibility of access?
- B. Should the District Court have expanded the “trivial showing” standard set forth in *Skidmore v. Led Zeppelin*, 952 F.3d 1051, 1068 (9th Cir. 2020)?
- C. Did the District Court err in failing to find an establishment of a chain of events linking Appellant’s Work and the Appellees’ work?
- D. Does Facebook’s widespread disposition of its users’ private data coupled with

the widespread accessibility of modern technological advancements such as artificial intelligence data scraping technology give rise to a reasonable possibility of access in the present case?

IV. STATEMENT OF THE CASE AND FACTUAL STATEMENT

A. Procedural History

This is an appeal of the District Court’s Order granting Appellees’ 12(b)(6) motion to dismiss Appellants’ copyright infringement claims with prejudice on the basis of a factual finding that Appellant failed to state a plausible theory of Defendant’s access to the Work. An overview of the relevant procedural history is as follows:

The Plaintiff-Appellant in case No. 2:20-cv-09843 filed her First Amended Complaint against Defendants-Appellees in the United States District Court for the Central District of California on February 22, 2021 (ER_214). On March 8, 2021, Defendants filed a Notice of Motion and Motion to Dismiss the case as to Defendant Perry (“First Motion”) (ER_215). On March 19, 2021, Plaintiff filed her Opposition to the First Motion (ER_215). On April 23, 2021, Defendants filed a Reply In Further Support of the First Motion (ER_217). On July 20 2021, the Court took the First Motion to amend under submission without oral argument (ER_219). On February 23, 2022, the Court issued an Order granting Defendants’ First Motion to dismiss the case as to Defendant Perry (ER_220). On August 31, 2022, the remaining Defendants filed a Notice of Motion and Motion to Dismiss the case

(“Second Motion”) (ER_222). On September 29, 2022, Plaintiff filed her Opposition to the Second Motion (ER_223). On October 7, 2022, the remaining Defendants filed a Reply in support of the Second Motion (ER_223). With respect to both aforementioned Motions, the Court found that Plaintiff failed “to allege sufficient facts to state a plausible theory of access to the lyrics of *Upgraded 2.0*, which is a necessary element of [Plaintiff’s claim.” (ER_030), and more specifically, that “Plaintiff failed to establish a chain of events linking the plaintiff’s work and the defendant’s access.” (ER_008-09).¹ On November 13, 2023, the Court granted the remaining Defendants’ Second Motion. Plaintiff-Appellant filed a Notice of the instant Appeal on December 12, 2023.

B. Statement of Facts

1. The Parties’ Works

In or about 2014, Appellant Michele Ronk created the unique and creative musical composition (lyrics) for a song she entitled *Upgraded 2.0* (ER_076 ¶ 18) (the “Work”). On September 23, 2014, Appellant shared her Work to her Facebook page via a status update containing an image of her desk and the text of the aforementioned lyrics (ER_077-78). On or about July 10, 2020, Appellee Hudson, a highly successful pop star professionally known as Katy Perry, released the master

¹ The Court also held that the Plaintiff’s work was not “widely disseminated” (ER_025) nor did the Plaintiff establish striking similarity between the Parties’ works in order to raise a presumption of access (ER_033). However, Appellant does not raise the foregoing two issues on Appeal.

recording, *Smile* (ER_083 ¶ 31; E.R_086 ¶ 41). Perry is signed to Appellee Capitol Records, LLC, which is owned by Appellee Universal Music Group (ER_083 ¶ 31). On July 11, 2020, Perry posted a picture to her Instagram page of a computer screen showing some of the lyrics from *Smile*; Appellant, who follows Perry on Instagram, immediately recognized prominent similarities between the *Smile* lyrics and that of the Work (ER_086 ¶ 41). Although the District Court found that the works were not strikingly similar as to raise a presumption of access, it did not reach the issue of substantial similarity (ER_033), which should be addressed on remand. Based on the foregoing allegations, the relevant time period regarding access is between the date Appellant originally posted the Work to Facebook on September 23, 2014 and the date Perry released *Smile* on or about July 10, 2020.

2. Facebook Shared its Users' Data with Partners

Unless otherwise supported by a citation to the evidentiary record, or a legal citation, the facts set forth below in Sections IV(B)(2)-(4) (which is also referenced in other sections in this brief) are provided by way of an “offer of proof” to demonstrate to the Court further facts that Appellant could allege in a Second Amended Complaint, as well as what facts discovery will likely reveal.

"Facebook operates one of the largest social media platforms in the world, with over one billion active users. About seven in ten adults in the United States use Facebook." *Campbell v. Facebook, Inc.*, 951 F.3d 1106, 1106 (2020) (*quoting Patel v. Facebook, Inc.*, 932 F.3d 1264, 1267 (9th Cir. 2019)). “Facebook collects

data from its users, including the types of content they access, the devices they use to access Facebook, their payment information, and their location. The collected data is used to individualize the content a user sees on Facebook.” *Sec. Litig. v. Facebook Inc.*, 87 F.4th 934, 941-42 (9th Cir. 2017). In addition, as detailed in a 2019 Federal Trade Commission (“FTC”) Complaint, Facebook also collects and stores other information about its users such as photos and videos, including any that a user has uploaded, as well as messages that a user posts. Complaint, Dkt. 1 ¶ 6(c), *In the Matter of Facebook, Inc.*, 19-cv-2184 (D.C. 2019). “Facebook has stored users’ profile information on a computer network that it controls” that can be accessed by third parties. *Id.* at ¶ 8.

The FTC Complaint alleged that, similar to Appellant’s post containing her Work, unbeknownst to users and in violation of Facebook’s representations in its privacy policy, “a user’s choice to restrict profile information to ‘Only Friends’ or ‘Friends of Friends’ would be ineffective as to certain third parties.” *Id.* at ¶ 14. “For example, a Platform Application with a narrow purpose . . . in many instances could access . . . every photo and video that the user had uploaded to Facebook’s website, despite the lack of relevance of this information to the Application.” *Id.* at ¶ 31. Moreover, as of the date of the FTC Complaint on July 24, 2019, “Facebook has collected and stored vast quantities of photos and videos that its users upload, including . . . more than 100 million photos and 415,000 videos from its users, collectively, every day.” *Id.* at ¶ 50. This information undoubtedly contains

Appellant's Work.

On July 24, 2019, the FTC secured a record-breaking \$5 billion civil penalty against Facebook for violating a 2012 Court order, which in part, barred Facebook from making misrepresentations about its data sharing activities. *See* Stip. Ord. for Civ. Penalty, Monetary Jdgmt., and Inj. Relief, Dkt. 2-1, at 3-4, *Facebook, supra* 19-cv-2184.² FTC Chairman Joe Simmons and Commissioners Noah Joshua Phillips and Christine S. Wilson issued a written public statement providing a summary of the FTC litigation against Facebook (*In re Facebook, Inc.*), detailing some of Facebook's data sharing activities, which gave rise to the FTC Complaint and civil penalty: "In 2012, Facebook entered into a consent order with the FTC, resolving allegations that the company misrepresented to consumers the extent of data sharing with third-party applications and the control consumers had over that sharing. The 2012 order barred such misrepresentations, and required Facebook to

² The 2012 FTC Order sought to regulate Facebook's representations as it relates to the information it obtains from and stores with respect to its users. The "Covered Information" as it is defined in the 2012 order, includes, but is not limited to "(a) a first or last name; (b) geolocation information sufficient to identify a street name and name of city or town; (c) an email address or other online contact information, such as an instant messaging User identifier or a screen name; (d) a mobile or other telephone number; (e) **photos and videos**; (f) Internet Protocol ("IP") address, User ID, or other persistent identifier that can be used to recognize a User over time and across different devices, websites or online services; (g) a Social Security number; (h) a driver's license or other government issued identification number; (i) financial account number; (j) credit or debit information; (k) date of birth; (l) biometric information; (m) any information combined with any of (a) through (l) above; or (n) Nonpublic User Information." *Id.*, at 11, *Facebook, supra*, 19-cv-2184.

establish a reasonable program to protect privacy. Our complaint announced today alleges that Facebook failed to live up to its commitments under that order.”³ FTC Commissioner Rebecca Kelly Slaughter echoed Facebook’s continued violations in her dissenting statement of July 24, 2019: “In April 2014, Facebook’s CEO and founder, Mark Zuckerberg, publicly assured Facebook users that third-party access to ‘friend’ data would stop. It did not.”⁴

In fact, that same year in 2014, Facebook “created a ‘reciprocity’ system in which certain third-party apps that provided ‘reciprocal value to Facebook’ could be ‘whitelisted,’ meaning that those apps were exempt from the ban on third-party data access and collection. *Sec. Litig. v. Facebook, Inc. (In re Facebook, Inc.)*, 87 F.4th 934, 945 (9th Cir. 2023). FTC Commissioner Rohit Chopra also issued a dissenting opinion on July 24, 2019, arguing that the July 2019 Order and penalties did not go far enough. In pertinent part, Chopra’s dissenting opinion stated “Facebook flagrantly violated the FTC’s 2012 order by deceiving its users and allowing pay-for-play harvesting by developers. Cambridge Analytica’s tactics of

³ Federal Trade Commission, *Statement of Chairman Joe Simons and Commissioners Noah Joshua Phillips and Christine S. Wilson Regarding the Matter of Facebook, Inc.*, at 1 (July 24, 2019), https://www.ftc.gov/system/files/documents/public_statements/1536946/092_3184_facebook_majority_statement_7-24-19.pdf (“FTC Majority Statement”).

⁴ Federal Trade Commission, *Dissenting Statement of Commissioner Rebecca Kelly Slaughter Regarding the Matter of Facebook, Inc.*, at 1 (July 24, 2019), https://www.ftc.gov/system/files/documents/public_statements/1536918/182_3109_slaughter_statement_on_facebook_7-24-19.pdf (“FTS Slaughter Dissent”).

profiling and targeting users were a small-scale reflection of Facebook’s own practices.”⁵ Chopra’s reference to Cambridge Analytica is a reference to the 2018 ‘data awakening,’ which occurred after news broke that Facebook had allowed the political consulting firm, Cambridge-Analytica, to access users’ personal data, which Cambridge Analytica improperly used to build tools aiding Donald Trump’s 2016 presidential campaign.⁶ This was the beginning of a series of reports that Facebook had given a myriad of third-party companies access to users’ personal data, including users’ private messages, names, contact details, and even the activity of friends through Facebook’s “integrated partnerships.” *Id.*

a. Facebook’s Demonstrated Data Sharing Activity

The Cambridge Analytica scandal laid bare Facebook’s “tactics of tricking users into sharing excessive amounts of personal data and then getting paid by third parties to target individual users.” (FTC Chopra Dissent, at 2). “This thirst for data has led the company to harvest intimate, personal details about tens of

⁵ Federal Trade Commission, *Dissenting Statement of Commissioner Rohit Chopra Regarding the Matter of Facebook, Inc.*, at 2 (July 24, 2019), https://www.ftc.gov/system/files/documents/public_statements/1536911/chopra_dissenting_statement_on_facebook_7-24-19.pdf (“FTC Chopra Dissent”).

⁶ *Facebook’s Data-Sharing Deals Exposed*, BBC (Dec. 19, 2018), <https://www.bbc.com/news/technology-46618582>. “According to the complaint against Cambridge Analytica filed July 24, 2020 (*In the matter of Kogan et al.*), Cambridge Analytica partnered with a Facebook application, the GSRApp, to access the Facebook platform and collect Facebook users’ profile data from approximately 250,000 Facebook users who directly accessed the app, as well as over 50 million of the direct users’ Facebook ‘friends.’” (FTC Slaughter Dissent, at 9).

millions of Americans on a scale and scope that are almost unimaginable. Facebook's data collection is both ongoing and increasing, as the company continues to add new means of surveillance that can be difficult to avoid. To facilitate further data acquisition, Facebook grants itself the right to surveil, own, and monetize users' private information by binding them to constantly evolving take-it-or-leave-it terms at sign-on." *Id.*

Reportedly, third parties with whom Facebook entered into data sharing agreements included major technology and entertainment corporations such as Amazon, Microsoft, Netflix, Pandora, Spotify, Sony, and Yandex. *Id.* "On June 3, 2018 . . . *The New York Times* reported that Facebook had continued sharing the date of users and their Facebook friends with dozens of whitelisted third parties like Apple, Microsoft, and Samsung with the users' express consent." *Sec. Litig. v. Facebook, Inc. (In re Facebook, Inc.)*, 87 F.4th 934, 945 (9th Cir. 2023). These 'whitelisted apps' were allegedly "allowed to completely override Facebook users' privacy settings." *Id.* For clarity, this practice of 'whitelisting' was demonstrably active from 2014 until at least 2018.

For example, Facebook allowed Netflix and Spotify to read Facebook users' private messages.⁷ *See also Campbell v. Facebook, Inc.*, 951 F.3d 1106, 1113 (9th

⁷ Gabriel J.X. Dance, et al., *As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants*, *The New York Times* (Dec. 18, 2018), <https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html> (ER_089).

Cir. 2020) (affirming class action settlement arising from allegations Facebook shared its users private Facebook messages with third parties, “enabling those third parties to generate customized content . . . on their own websites informed by this data”).⁸

Facebook also reportedly allowed Yahoo to view streams of friends’ posts, despite public statements that it had stopped that type of sharing years earlier. *Id.* Furthermore, former employees of Facebook reported that partnerships were often exempt from Facebook’s data policy. *Id.* Records of deals between Facebook and over 150 companies dating back to 2010 show these companies sought access to the data of hundreds of millions of people each month. *Id.* In 2018 “Facebook did say that it had mismanaged some of its partnerships, allowing certain companies’ access to continue long after they had shut down the features that required the data.” *Id.* Not only did Facebook knowingly give wide access to its partners, but in some instances software bugs exposed millions of user photos to 1500 third-party applications. Tony Romm, *Facebook says a new bug allowed apps to access private photos of up to 6.8 million users*, The Washington Post (Dec. 14, 2018)⁹

⁸ Facebook claimed it ceased disclosing private message data to third parties in 2014; however, Facebook acknowledged that as of 2020, “Facebook was actively accessing private messages . . . [and] Facebook’s ongoing retention of the data collected from private messages meant there was a risk that it would . . . resume sharing the data with third parties, or begin using the data for some other purpose.” *Id.* at 1119.

⁹ Tony Romm, *Facebook says a new bug allowed apps to access private photos of up to 6.8 million users*, The Washington Post (Dec. 14, 2018)

<https://www.washingtonpost.com/technology/2018/12/14/facebook-says-new-bug->

3. Facebook and UMG are Strategic Partners since 2017

In 2017, UMG and Facebook joined forces to create a first-of-its-kind partnership in a multi-year agreement (ER_087 ¶ 44).¹⁰ Under the partnership, UMG alleges that it has licensed its recorded music and publishing catalogs for video and “other social experiences” on Facebook, Instagram, and Oculus. *Id.* This allows users to upload videos containing licensed music and personalize their music experiences. *Id.* Although neither Facebook nor UMG have discussed exactly what data UMG might have access to (*Id.*), pursuant to Facebook’s demonstrated ‘whitelisting’ practices and data sharing with its corporate partners, this ‘integrated partnership’ likely gives UMG access to a host of Facebook user data. Facebook and certain third-party recipients of Facebook’s users’ data have claimed at times they were unaware or oblivious to the data to which they were unknowingly provided access; however, the Court should bear in mind that big data is big business, and Facebook has wielded the mammoth amount of user data in its coffers to line those coffers. “From the time of the original 2012 Facebook order to 2018, Facebook’s gross annual revenue increased more than 1000% from

[allowed-apps-access-private-photos-up-million-users/](#).

¹⁰ *Facebook and Universal Music Group Strike Unprecedented Global Agreement*, UMG (Dec. 21, 2017) <https://www.universalmusic.com/facebook-universal-music-group-strike-unprecedented-global-agreement/> (ER_087, fn. 10). In 2018, Sony and Warner followed UMG in signing similar agreements with Facebook. Motley Fool, *Facebook Scores Deal With Last Major Record Label*, Fox Business (March 9, 2018), <https://www.fool.com/investing/2018/03/09/facebook-scores-deal-with-last-major-record-label.aspx>

\$5 billion to over \$55 billion.” (FTC Slaughter Dissent, at 8). As the ‘whitelisting’ practice was in effect from at least 2014 to 2018, the trafficking of Facebook’s users’ data is clearly a driving factor of the revenue boom.¹¹

In addition, BBC reports that Facebook “reward[s] companies with access to data that others are denied, if they place a high value on the business they do together. This is just another form of selling.”¹² A USPTO report regarding artificial intelligence also echoes the value of big data: “Access to data for initial development and ongoing training is necessary for AI development. This means that data and datasets, including their collection and compiling, have value, particularly ‘big data’ (i.e. extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations).” USPTO, *Public Views on Artificial Intelligence and Intellectual Property Policy*, at 15 (October 2020).¹³ In other words, the data is extremely valuable, and both Facebook and its

¹¹ “The price of Facebook’s stock declined significantly in the week that followed the *Cambridge Analytica* revelations,” dropping roughly 18%, or \$100 billion in market capitalization in one week. *Sec. Litig. v. Facebook, Inc. (In re Facebook, Inc.)*, 87 F.4th 934, 944-45 (9th Cir. 2023). On July 25, 2018, Facebook announced unexpectedly low revenue growth in a Q2 earnings call, which it attributed to “putting privacy first” and implementing the European Union’s General Data Protection Regulation. *Id.* at 945. “The day after the earnings call, Facebook’s stock price dropped nearly 19%.” *Id.* Clearly, any souring on its ability to monetize its users’ data is a meaningful hit to Facebook’s pockets, which is demonstrative of the value of said data to Facebook and its partners.

¹² *Facebook’s Data-Sharing Deals Exposed*, BBC (Dec. 19, 2018),

<https://www.bbc.com/news/technology-46618582>

¹³ https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf

partners, including UMG know it.

In fact, depending on the third-party's purposes, and the availability of artificial intelligence programs and other data scraping tools, the value of Facebook's data is that it could be used for virtually *anything*. An executive at the World Privacy Forum warned that Facebook loses control of the data once it is shared and then "[i]t could be customized. It could be fed into an algorithm and decisions could be made about you based on that data."¹⁴ Another expert warned, "[i]f Cambridge Analytica was able to acquire information on tens of millions of Facebook users so quickly and easily, and then keep the information for years without Facebook suspecting otherwise, then that shows a serious flaw in Facebook's ability to keep exclusive control over its information." *Sec. Litig. v. Facebook, Inc. (In re Facebook, Inc.)*, 87 F.4th 934, 945 (9th Cir. 2023). Assuming UMG has access to Facebook user data, modern algorithms and artificial intelligence programs allow Defendants to quickly search and locate specific information in a way that makes access easier than ever before.

4. Availability of Artificial Intelligence Programs

Plaintiff alleged in her FAC that Defendants may have been able to use SQL queries to obtain with ease any particular data set (including Appellant's Work) with accuracy and ease (ER_093 ¶ 50 – ER_094 ¶ 56). In addition, the evolution of

¹⁴ Dance, et al., *supra* note 7 (ER_089, fn. 16).

Artificial Intelligence (AI) programs, particularly in the realm of natural language processing (NLP), has witnessed remarkable advancements in recent years, which may also have been utilized towards the same end. Before the release of *Smile* in July 2020, NLP models had revolutionized the ability to understand and generate human-like text, enabling advancements in machine translation, sentiment analysis, question-answering systems, and more. As AI practitioners utilize large datasets, such as those provided by Facebook to its partners (including UMG), the possibility of unknowingly creating derivative works is significant. A brief overview of just a few AI programs that were readily accessible prior to the release of *Smile* in July 2020 are as follows:

a. GPT-1

GPT-1, or the “Generative Pre-trained Transformer,” is based on the transformer architecture, which is a neural network architecture designed for NLP tasks.¹⁵ This architecture was introduced by Vaswani et al. in the paper “Attention is All You Need.”¹⁶ It utilizes a self-attention mechanism that allows the model to weigh the importance of each word in a sequence concerning every other word in the sequence. By doing so, the model can capture long-range dependencies and

¹⁵ Partha Pratim Ray, *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope*, ScienceDirect, (Apr. 14, 2023), <https://www.sciencedirect.com/science/article/pii/S266734522300024X#bib152>.

¹⁶ Ashish Vaswani et al., *Attention Is All You Need*, arXiv, (Jun. 12, 2017), <https://arxiv.org/pdf/1706.03762.pdf>.

relationships between words across an entire dataset efficiently. GPT-1 has 117 million parameters, which are the variables that the model learns during training.¹⁷ These parameters capture the relationships between words in the text data and enable the model to make predictions about sequences of text.

GPT-1 was pre-trained on a large corpus of text data, which includes books, articles, and web pages. During pre-training, the model is trained to predict the next word in a sequence given the previous words.¹⁸ This process exposes the model to a diverse range of language patterns, styles, and contexts present in a dataset. As a result, GPT-1 becomes proficient at understanding and generating coherent text. After pre-training, GPT-1 can be fine-tuned on specific downstream tasks, such as language translation, sentiment analysis, or text classification.¹⁹ Fine-tuning involves providing the model with labeled datasets for the target task and updating its parameters to optimize performance on that task. For example, to fine-tune GPT-1 for sentiment analysis, you would provide it with a labeled dataset of text data along with their corresponding sentiments (positive, negative, neutral) and train it to predict the sentiment of a given text input.²⁰

GPT-1 achieved impressive results on a wide range of NLP tasks, demonstrating the effectiveness of pre-training on large amounts of text data for

¹⁷ Ray, *supra* note 15.

¹⁸ *Id.*

¹⁹ *Id.*

²⁰ *Id.*

improving language understanding. By pre-training on diverse text data and fine-tuning on specific tasks, GPT-1 can generalize well to various NLP tasks without requiring task-specific architectures or extensive labeled datasets.²¹ GPT-1 also has the ability, like other models based on the transformer architecture, of deciphering large datasets. Its self-attention mechanism is crucial for this task because it enables the model to understand the context and meaning of words within the broader context of the dataset.²²

b. GPT-2

GPT-2, like its predecessor GPT-1, is based on the transformer architecture. However, GPT-2 is more powerful, with 1.5 billion parameters compared to GPT-1's 117 million.²³ This increased parameter count allows GPT-2 to capture more intricate patterns and relationships in the text data it processes. Additionally, this transformer architecture is highly scalable, allowing GPT-2 to efficiently process and understand large amounts of text data by leveraging self-attention mechanisms.

The massive corpus of text data GPT-2 is pre-trained on is much larger than GPT-1's, enabling GPT-2 to learn more diverse language patterns and styles. A large dataset called WebText containing over 40GB of web pages filtered from outbound links on Reddit was also used in pre-training.²⁴ This dataset provides

²¹ *Id.*

²² Vaswani, *supra* note 16.

²³ Ray, *supra* note 15.

²⁴ *Id.*

GPT-2 with a diverse range of text data from various sources on the internet, allowing it to learn from a wide range of language styles and topics. It is also trained to predict the next word in a sequence given the previous words, generating coherent and contextually relevant text.²⁵

GPT-2 demonstrates impressive generalization capability, meaning that it can effectively apply its learned knowledge to new tasks and domains. This ability is crucial for deciphering large datasets because it allows GPT-2 to adapt and perform well on diverse NLP tasks without requiring task-specific fine-tuning.²⁶ By leveraging its pre-trained knowledge and understanding of language, GPT-2 can process and interpret large datasets across various applications, such as text classification, summarization, and question-answering.

c. Bidirectional Encoder Representations from Transformers (BERT)

BERT was developed by Google and introduced in a paper by Jacob Devlin et al. in 2018.²⁷ It represents a significant advancement in NLP by using bidirectional training, allowing it to understand the context of words from both directions within a sentence.²⁸ This bidirectional approach is made possible by the

²⁵ *Id.*

²⁶ *Id.*

²⁷ Jacob Devlin et al., *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, arXiv, (Oct. 11, 2018), <https://arxiv.org/pdf/1810.04805.pdf>.

²⁸ Ray, *supra* note 15.

highly scalable transformer architecture, which forms the backbone of BERT.

BERT is pre-trained on two unsupervised tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP).²⁹ In MLM, random words in a sentence are replaced with a special token, and the model is trained to predict the original words based on the context provided by the surrounding words. This task enables BERT to learn deep contextual representations of words within sentences. In NSP, BERT learns to predict whether pairs of sentences are connected in a logical sequence. By pre-training on such diverse tasks and text data, BERT learns rich contextual representations of words and sentences, enabling it to decipher large datasets with varied language patterns and styles.

After training, BERT can be fine-tuned on downstream tasks such as text classification, named entity recognition, and question-answering.³⁰ Fine-tuning involves providing BERT with labeled data for the specific task and updating its parameters to optimize performance. Because BERT has already learned rich contextual representations of language during pre-training, fine-tuning allows it to further refine its understanding of the data and optimize its performance, making it well-suited for handling diverse NLP tasks on large datasets.

d. Text-to-Text Transfer Transformer (T5)

T5 is an NLP model developed by Google, designed to handle a wide range

²⁹ *Id.*

³⁰ Devlin et al., *supra* note 27.

of NLP tasks.³¹ It represents a significant advancement in NLP by adopting a text-to-text approach, simplifying the task-specific fine-tuning process and allowing for improved performance across various applications.³² T5 is based on the transformer architecture, which has been highly influential in the NLP domain due to its exceptional performance in capturing contextual relationships in text data.

T5 is trained using a text-to-text approach where input and output are both represented as text. This approach simplifies the process of fine-tuning the model for specific tasks by converting them into a common format.³³ By training the model to generate text outputs from text inputs, T5 learns to perform a wide range of NLP tasks, including text summarization, language translation, and question-answering. This unified training framework promotes transfer learning, enabling T5 to use knowledge learned from one task to improve performance on others.³⁴

T5 is pre-trained on a large-scale dataset called Colossal Clean Crawled Corpus (C4), which is a cleaned and deduplicated version of the Common Crawl dataset.³⁵ This extensive pre-training helps T5 learn general language understanding capabilities by exposing it to diverse linguistic patterns and structures. Subsequently, T5 can be fine-tuned for specific tasks by providing task-

³¹ Colin Raffel et al., *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, arXiv, (Oct. 23, 2019), <https://arxiv.org/pdf/1910.10683.pdf>.

³² Ray, *supra* note 15.

³³ *Id.*

³⁴ *Id.*

³⁵ *Id.*

specific datasets and adjusting its parameters accordingly. Fine-tuning allows T5 to adapt its learned representations to the nuances of the target task, achieving superior performance compared to models trained from scratch.³⁶

T5's ability to decipher large datasets stems from its text-to-text approach, transformer architecture, and extensive pre-training on a large-scale corpus of text data. By training on the C4, T5 learns rich representations of language structures and semantics, enabling it to understand and generate coherent text across diverse domains.³⁷ Additionally, T5's fine-tuning capability allows it to adapt quickly to new tasks and datasets, further enhancing its effectiveness in deciphering large volumes of text data.

e. XLNet

XLNet is an auto-regressive language model developed to combine the strengths of both Transformer-XL and BERT.³⁸ It represents a significant advancement in NLP by addressing some limitations of other transformer models like BERT and GPT. XLNet's architecture and training methodology enable it to capture long-range dependencies within text effectively.³⁹ These dependencies capture how earlier parts of a sentence or a text influence the interpretation of later

³⁶ *Id.*

³⁷ *Id.*

³⁸ *Id.*

³⁹ Zhilin Yang et al., *XLNet: Generalized Autoregressive Pretraining for Language Understanding*, arXiv, (Jun. 19, 2019), <https://arxiv.org/pdf/1906.08237.pdf>.

parts, even when separated by many intervening words, thus making XLNet highly effective for a wide range of NLP tasks.

XLNet's training objective combines both auto-regressive and auto-encoding language modeling approaches.⁴⁰ In the auto-regressive approach, XLNet predicts the next token in a sequence based on the context of previous tokens, similar to models like GPT. In the auto-encoding approach, XLNet predicts masked tokens in a given sequence by considering both left and right contexts, similar to BERT's masked language model objective. These approaches allow XLNet to produce coherent and contextually relevant text across a wide range of tasks, including language modeling, text classification, and machine translation.⁴¹

Additionally, XLNet employs a permutation-based training approach where it learns to predict a token based on a random arrangement of the input sequence.⁴² This approach allows XLNet to capture bidirectional context like BERT while avoiding some of the issues associated with BERT's masked language model objective.⁴³ By randomly permuting the input sequence during training, XLNet is exposed to various arrangements of the context, enabling it to learn rich representations of bidirectional dependencies within text data.

/ / /

⁴⁰ Ray, *supra* note 15.

⁴¹ Yang et al., *supra* note 39.

⁴² *Id.*

⁴³ Ray, *supra* note 15.

f. Robustly Optimized BERT Pretraining Approach (RoBERTa)

RoBERTa is a variant of BERT developed by Facebook AI.⁴⁴ It represents an enhanced version of the BERT model, building on BERT's architecture by introducing a series of optimizations and training improvements to achieve better performance and accuracy in NLP tasks.⁴⁵ RoBERTa incorporates several improvements to the pre-training process compared to BERT. These improvements include the use of larger batch sizes, longer training times, and optimized hyperparameters.⁴⁶ By training on larger batches, removing the next sentence prediction task, using longer sequences, and training on more data, RoBERTa achieves a more accurate and well-rounded language model.

One of the key enhancements of RoBERTa is its focus on robust pre-training. It is designed to be more enduring to noise and variance in training data, which helps improve its performance on downstream tasks.⁴⁷ By training on this larger dataset, RoBERTa learns more robust representations of language, allowing it to generalize better to unseen data and perform more reliably across various NLP tasks. RoBERTa maintains the highly flexible architecture of BERT, allowing it to be fine-tuned for a wide range of NLP tasks.⁴⁸ Whether it is text classification,

⁴⁴ *Id.*

⁴⁵ Yinhan Liu et al., *RoBERTa: A Robustly Optimized BERT Pretraining Approach*, arXiv, (Jul. 26, 2019), <https://arxiv.org/pdf/1907.11692.pdf>.

⁴⁶ Ray, *supra* note 15.

⁴⁷ *Id.*

⁴⁸ *Id.*

named entity recognition, sentiment analysis, or machine translation, RoBERTa's flexible architecture enables it to adapt to different tasks and datasets with ease. This flexibility makes RoBERTa a versatile tool for NLP researchers and practitioners, capable of addressing diverse challenges in text processing and analysis.

g. DistilBERT

DistilBERT, a variant of the BERT model, was developed by Hugging Face with the intention of retaining the high performance of BERT while reducing its computational complexity and memory footprint.⁴⁹ Its architecture closely resembles that of BERT however, it incorporates techniques such as knowledge distillation to compress the original BERT model into a more lightweight version without compromising its effectiveness in NLP tasks.⁵⁰

During pre-training, DistilBERT learns to mimic the behavior of the larger BERT model by distilling the knowledge it has acquired. This process involves training DistilBERT to predict the same outputs as BERT while using fewer parameters and computational resources.⁵¹ By distilling the knowledge from BERT, DistilBERT can achieve similar performance levels while being more efficient. In addition to knowledge distillation, DistilBERT incorporates various

⁴⁹ Victor Sanh et al., *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*, arXiv, (Oct. 2, 2019), <https://arxiv.org/pdf/1910.01108.pdf>.

⁵⁰ *Id.*

⁵¹ *Id.*

compression techniques to reduce its size and computational requirements.⁵²

Despite its smaller size and reduced complexity, DistilBERT maintains its versatility and adaptability across a wide range of NLP applications. Its efficient training process and optimized architecture enable it to deliver reliable performance in real-world scenarios, while its ease of deployment makes it accessible to a broader audience of developers and researchers.⁵³ As a result, DistilBERT emerges as a valuable asset in the field of NLP, offering a balance between performance and efficiency that caters to the diverse needs of the NLP community.

h. ALBERT

ALBERT (A Lite BERT) is a variant of the BERT model developed by Google Research and Toyota Technological Institute at Chicago to address some of the limitations of the original BERT model, particularly in terms of its computational cost and memory requirements.⁵⁴ It is designed to be a more efficient and scalable version of BERT while maintaining or even improving its performance by training it on larger datasets to deploy it more efficiently in real-world applications.⁵⁵

⁵² *Id.*

⁵³ *Id.*

⁵⁴ Zhenzhong Lan et al., *ALBERT: A Lite BERT for Self-supervised Learning of Language Representations*, arXiv, (Sep. 26, 2019), <https://arxiv.org/pdf/1909.11942.pdf>.

⁵⁵ *Id.*

ALBERT is trained using the same pre-training objectives as BERT, such as masked language modeling and next sentence prediction.⁵⁶ It follows the same transformer architecture as BERT as well, however, ALBERT incorporates several modifications and optimizations to improve efficiency and scalability.⁵⁷

One of the key innovations of ALBERT is the use of parameter sharing across layers.⁵⁸ In traditional transformer models like BERT, each layer has its own set of parameters, leading to redundancy and increased memory requirements. In ALBERT, parameters are shared across layers, reducing the overall number of parameters and making the model more efficient.⁵⁹ ALBERT also introduces a technique called “factorized embedding parameterization,” which further reduces the memory footprint of the model.⁶⁰ This technique decomposes the embedding matrix into smaller matrices, resulting for more efficient storage and computation.

ALBERT can be used as a drop-in replacement for BERT in various NLP applications where efficiency and scalability are a concern. Its smaller size and reduced memory requirements make it particularly suitable for deployment in resource-constrained environments such as mobile devices, edge computing devices, and real-time applications.⁶¹

⁵⁶ *Id.*

⁵⁷ *Id.*

⁵⁸ *Id.*

⁵⁹ *Id.*

⁶⁰ *Id.*

⁶¹ *Id.*

i. Stanford CoreNLP

Stanford CoreNLP is an NLP toolkit developed by the Stanford Natural Language Processing Group, aiming to provide comprehensive tools and algorithms for text analysis.⁶² Its development was driven by the need for a versatile and easy-to-use toolkit that could handle a variety of NLP tasks. Built on Java, it utilizes a modular architecture, enabling users to employ separate modules for different tasks, such as tokenization, part-of-speech tagging, named entity recognition, dependency parsing, sentiment analysis, and coreference resolution.⁶³

Key features of Stanford CoreNLP include its ability to tokenize text, break it down into individual words or tokens, and assign grammatical tags through part-of-speech tagging.⁶⁴ Named entity recognition identifies and classifies entities such as people, organizations, and locations. Dependency parsing analyzes the arrangement and organization of different parts of speech in sentences to determine relationships between words. Additionally, sentiment analysis gauges the emotional tone of text, while coreference resolution identifies and resolves references to the same entity across various parts of a text.

Stanford CoreNLP is accessible through both local installation and a web

⁶² Christopher D. Manning et al., *The Stanford CoreNLP Natural Language Processing Toolkit*, The Stanford Natural Language Processing Group, (2014), <https://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf>.

⁶³ *Id.*

⁶⁴ *Id.*

service provided by the Stanford NLP Group, making it convenient for developers to obtain and use. Moreover, the toolkit comes with pre-trained models for various languages and tasks, reducing the need for extensive training data.⁶⁵ Researchers and developers widely employ Stanford CoreNLP for both practical applications and research purposes in NLP, benefiting from its versatility, reliability, and actively developing community.

j. Availability of Artificial Intelligence Programs Prior to July 2020

Prior to July 2020, accessibility to AI programs, particularly those in the field of NLP, was on the rise. Various factors contributed to this accessibility, including the emergence of open-source libraries and models. Leading AI models such as GPT-2, BERT, and RoBERTa were released as open-source projects, allowing researchers, developers, and students to access and experiment with state-of-the-art models without needing to build them from scratch.⁶⁶

This accessibility was further enhanced by the availability of pre-trained models for download and use in various tasks. These models reduced the barrier to entry for individuals lacking extensive computational resources or large datasets for training. Additionally, user-friendly interfaces provided by platforms like Hugging Face and other AI-focused websites made it easier for developers to

⁶⁵ *Id.*

⁶⁶ Ray, *supra* note 15.

implement NLP capabilities in their applications.⁶⁷

Hugging Face, a prominent AI research company and open-source community, played a significant role in facilitating this new accessibility. Founded in 2016 by Clément Delangue, Julien Chaumont, and Thomas Wolf, Hugging Face initially started as a chatbot developer but quickly transitioned to focus on creating an accessible and collaborative platform for machine learning and NLP.⁶⁸

The company's main offering is its Transformer Library, an open-source library that provides easy access to state-of-the-art NLP models such as BERT, GPT, and RoBERTa.⁶⁹ Users can load pre-trained models, fine-tune them for specific tasks, and use them for inference. Additionally, Hugging Face offers a Model Hub, a central repository where users can find and share pre-trained models for various NLP tasks.⁷⁰ This hub fosters collaboration and resource sharing by allowing users to upload and share their own models. Hugging Face also provides a Datasets Library, granting access to a vast collection of publicly available datasets.⁷¹ This library enables users to easily load, process, and manipulate data

⁶⁷ *Hugging Face – The AI Community Building the Future*, HUGGING FACE, <https://huggingface.co/> (last visited Apr. 19, 2024).

⁶⁸ Ben Lutkevich, *What is Hugging Face?*, TECHTARGET, <https://www.techtargget.com/whatis/definition/Hugging-Face> (last updated Sept. 2023).

⁶⁹ *Transformers*, HUGGING FACE, <https://huggingface.co/docs/transformers/en/index> (last visited Apr. 19, 2024).

⁷⁰ *Models*, HUGGING FACE, <https://huggingface.co/models> (last visited Apr. 20, 2024).

⁷¹ *Datasets*, HUGGING FACE, <https://huggingface.co/datasets> (last visited Apr. 20,

for training and testing NLP models. The platform promotes community engagement by encouraging contributions and collaboration. Users can interact with other researchers and developers, share models, and contribute to open-source projects.⁷² Hugging Face supports this with comprehensive documentation and tutorials to guide users through the process of using its tools.⁷³

People can use Hugging Face in various ways depending on their goals and expertise. For beginners, pipelines (high-level application programming interfaces) offer an easy way to perform NLP tasks with minimal coding, allowing users to experiment with sentiment analysis, translation, and text generation using pre-trained models.⁷⁴ More advanced users can fine-tune pre-trained models on custom datasets to adapt them to specific use cases. Developers can use the Transformers and Datasets libraries to build custom models from scratch, allowing for greater flexibility in designing models tailored to specific tasks and data.

Overall, Hugging Face is a versatile platform that caters to a wide range of NLP practitioners, from beginners to experts. Its open-source approach, community-driven model hub, and user-friendly tools make it a valuable resource for anyone interested in NLP and machine learning. This aligns well with the

2024).

⁷² *Posts*, HUGGING FACE, <https://huggingface.co/posts> (last visited Apr. 19, 2024).

⁷³ *Transformers*, *supra* note 69.

⁷⁴ *Pipelines for Inference*, HUGGING FACE, https://huggingface.co/docs/transformers/pipeline_tutorial (last visited Apr. 20, 2024).

broader trend of increasing accessibility to AI programs and resources prior to July 2020.

5. Summary of Facts and Alleged Chain of Events Between Appellant's Work and Appellees' Access

Appellant posted her Work to her Facebook page on September 23, 2014 (ER_077 ¶ 20 – ER_078) and Appellee released her derivative song *Smile* on July 10, 2020 (ER_083 ¶ 31; E.R_086 ¶ 41). Between that time period, there is more than enough demonstrable evidence to infer that (1) Facebook shared its users' data (including Appellant's Work) to its strategic partners (*see supra* Section IV(B)(2)); (2) UMG was one of Facebook's highly publicized and important strategic partners since at least 2017 (ER_044, 049) such that it is likely UMG acquired access to Facebook's large data sets (including Appellant's Work) (*see supra* Section IV(B)(3)); (3) pursuant to Facebook's history of whitelisting, which was admittedly in effect during the time period of its publicized UMG-partnership, both Facebook UMG likely intended to derive monetary benefit from the sharing of the data (ER_046-49) (*see supra* Section IV(B)(2)-(3)); (4) As Facebook's data may be used for a multitude of purposes from content generation to targeted advertising (to name a few uses), it is reasonable that UMG made that data available to its subsidiaries, including but not limited to Appellee Capitol Records for its music creation and advertising purposes (*see supra* Section IV(B)(3)); and (5), it is reasonable that as did the general public, UMG had access to and likely utilized one or more artificial

intelligence programs (or other data scraping technologies) to train upon, assess, learn, or otherwise incorporate information from the Facebook data in order to create derivative works, such as *Smile* (*see supra* Section IV(B)(4)).

The foregoing is by no means an exhaustive recitation of the facts; moreover, given that the Central District’s dismissal of the action results from a *factual* determination that Plaintiff failed to plead adequate facts to show a “reasonable” possibility of access, this Court should reverse and remand the case, and permit the Parties to engage in discovery regarding the somewhat complex, yet highly practical and realistic chain of events as plead by Appellant and set forth above with greater particularity. Further, Appellant’s additional facts submitted herein as an offer proof demonstrate that, at a minimum, leave to file a Second Amended Complaint and discovery is warranted in the present situation. The current digital landscape of social media (specifically Facebook) data sharing, coupled with the ubiquitous ability for *anyone* to access and use artificial intelligence software applications to digest and incorporate copyrighted subject matter into an ostensibly ‘original’ or ‘independently created’ product necessarily requires some level of discovery to resolve questions of access.

V. SUMMARY OF THE ARGUMENT

This Court should either reverse the grant of Appellees’ motions to dismiss or grant but permit Appellant to file a Second Amended Complaint because this Court should consider key facts as set forth above regarding the reasonability of

access under Appellant’s chain of events theory. First, the Court should consider whether the facts set forth above give rise to a plausible access argument. Second, the Court should consider whether, given the state of data collection by Facebook, the documented sharing thereof, and the use of artificial intelligence programs, an expansion of the ‘trivial access’ standard as set forth in *Skidmore v. Led Zeppelin* is warranted. Third, the Court should consider the dangerous implication of the use of artificial intelligence programs by third parties upon a privately (and secretly) obtained data set such as Facebook’s users’ content.

VI. ARGUMENT

A. STANDARD OF REVIEW

The appellate court reviews dismissals under Federal Rule of Civil Procedure 12(b)(6) *de novo*, accepting the plaintiff’s allegations as true and construing them in the light most favorable to the plaintiff. *Zucco Partners, LLC v. Digimarc Corp.*, 552 F.3d 981, 989 (9th Cir. 2009). At this stage of litigation, a complaint’s factual allegations need not be detailed. *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 555 (2007). They “must be enough to raise a right to relief above the speculative level” and to “state a claim to relief that is plausible on its face.” *Id.* at 555, 570. To state a claim for copyright infringement, a plaintiff must plausibly allege two things: 1) ownership of a valid copyright, and (2) copying of protected elements of the work that are original. *Shaw v. Lindheim*, 919 F.2d 1353, 1556 (9th Cir. 1990) (“a plaintiff may establish copying by showing that the infringer had

access to the work and that the two works are substantially similar).

On appeal, denial of a motion for leave to amend is reviewed for abuse of discretion, “strictly review[ing] such denial in light of the strong policy permitting amendment.” *Amerisource Bergen Corp. v. Dialysist W., Inc.*, 465 F.3d 946, 949, 957 (9th Cir. 2006). Here, Appellants have alleged facts that have “raise[d] a right to relief above the speculative level” and “is plausible on its face.” Thus, this Court should reverse the dismissal or require the lower court to grant leave to amend and then re-analyze the case applying the correct test for substantial similarity, and the other recognized doctrines set forth herein.

B. GIVEN THE CURRENT TECHNOLOGICAL LANDSCAPE, THE COURT SHOULD EXPAND THE CIRCUMSTANCES WHERE THERE IS A ‘REASONABLE POSSIBILITY OF ACCESS’

This case involves modern practices that can only be addressed by a modern interpretation of the law. Judicial interpretation of copyright law as it stands today has not kept pace with the rapidly changing technological landscape, which unknowingly provides a safe haven for sophisticated infringers, like Appellees, to inconspicuously avoid liability. Plaintiffs are detrimentally limited, at the pleadings stage, to making allegations about elaborate, inter-firm transactional and data analytic schemes of the corporate Appellees without the aid of the discovery process. It is clear that the District Court did not fail to state the reasoning and basis for its decision. However, this Court can and should be empowered to reconsider how the lower court’s ruling was based on an inadequate interpretation

of the facts and the law, and to make its own determination as to its correct application.

De novo review of the district court's grant of a motion to dismiss must accept all factual allegations in the complaint as true, and construe them in the light most favorable to Appellant. *Narayanan v. British Airways*, 747 F.3d 1125, 1127 (9th Cir. 2014). In doing so, this Court can conduct its own legal analysis to arrive at the appropriate conclusion, with no deference to the trial court's ruling.

Appellate Practice in Federal and State Courts § 3.05. Appellate courts are well within their power to set aside a district court's findings which rest on an erroneous view of the law. *Pullman-Standard v. Swint*, 456 U.S. 273, 287 (1982).

Accordingly, a lower court's conclusions on questions of the application, interpretation, and construction of law are subject to an appellate court's independent review. 4 *Federal Litigation Guide* § 48.123 (2024). Decisions regarding access are quickly becoming obsolete. In reconsidering the merits of Appellant's pleadings, this Court should evaluate how the current state of the law on access, as constructed, is lagging behind the common technological practices of firms who discretely amass and exploit vast amounts of data, shielded from legal action by those whose work is unassumingly swallowed up in the process.

Appellant's position is not newly-raised on appeal. The question of access, and the adequacy of its current legal interpretation, are already central to the District Court's ruling. Nevertheless, Appellate Courts have the discretion, on a

case-by-case basis, to decide what questions may be taken up and resolved for the first time on appeal. *Singleton v. Wulff*, 428 U.S. 106, 120 (1976). This Court exercises that discretion to prevent a miscarriage of justice” *Baccei v. United States*, 632 F.3d 1140, 1149 (9th Cir. 2011) (citing *Kimes v. Stone*, 84 F.3d 1121, 1126 (9th Cir. 1996)). This Court’s discretion stops short of reframing an appeal “to review what would be in effect a different case than the one decided by the district court.” *Baccei, supra* 632 F.3d at 1149. These principles are not a bar to reanalyzing issues which have been raised below, modifying them, or using new legal theories to support an appellant’s position, as is the case here. A good faith consideration of Appellant’s arguments, that the law on access should be expanded in order to keep up with the technological innovations of corporations, is not a reframing of the issues that the District Court decided below. On the contrary, there is a direct link between the decision of the lower Court and the issues presented in this brief.

This case is a testament to the reality that the law often fails to keep up with technology. Appellant firmly believes this case provides a reasonable basis in law and fact to extend the law on the question of access to its next logical place. “A complaint should not be dismissed unless it appears beyond doubt that plaintiff can prove no set of facts in support of his claim which would entitle him to relief.”

Parks Sch. of Business, Inc. v. Symington, 51 F.3d 1480, 1484 (9th Cir. 1995).

Appellant’s pleadings and the well-supported facts stated hereinabove make a good

faith showing of a chain of events that materially implicate Appellees in a web of dealings that are capable of discrete, but nonetheless, incredibly damaging infringement practices. It is the current strictures of the access prong that are precluding Appellant from vindicating her rights. The present approach to access substantially prejudices similarly situated plaintiffs whose livelihoods depend on the lawful exploitation of their work. This Court has the opportunity to level the playing field by expanding the parameters of the access prong to where they should already be.

C. THE COURT SHOULD RECONSIDER WHAT ‘REASONABLE POSSIBILITY’ OF ACCESS MEANS IN LIGHT OF MODERN DATA SHARING AND THE AVAILABILITY OF ARTIFICIAL INTELLIGENCE PROGRAMS

In expanding the judicial interpretation of what constitutes a “reasonable possibility” of access, the Court should consider the application of the “trivial” showing standard contemplated in the 2020 landmark decision, *Skidmore v. Led Zeppelin*. *Skidmore v. Led Zeppelin*, 952 F.3d 1051, 1068 (9th Cir. 2020). The *Skidmore* court stated that while wide dissemination of the copyrighted work is often how access is proved, “[a]s a practical matter, the concept of ‘access’ is increasingly diluted in our digitally interconnected world.” *Id.* The court then explained that due to the “ubiquity of ways to access media online,” a plaintiff can establish access through a “trivial showing that the work is available on demand.” *Id.* The acknowledgement of the “trivial showing” standard by the Ninth Circuit

reflects a pragmatic approach to address the challenges posed by the digital age in proving access. By recognizing the ubiquity of online media access and dilution of the traditional concept of access, courts have appeared to open the door for an adaptation of precedent to align with the realities of modern technology. Fully realizing this evolving perspective would not only facilitate but also streamline the process for copyright holders to readily establish access to their works.

The *Skidmore* standard has been cited in multiple cases since its establishment. In 2020, the Central District of California explained that access can be established using a “trivial showing that the work is available on demand.” *Gray v. Perry*, No. 2:15-CV-05642-CAS-JCx, 2020 U.S. Dist. LEXIS 46313, at *42 (C.D. Cal. Mar. 16, 2020) (quoting *Skidmore*, 952 F.3d at 1068). “The Court must defer to the jury's determination of reasonableness on this question. Even where a plaintiff's theory of access is ‘fantastic,’ the “plaintiffs credibility, even as to those improbabilities, should be left to the jury.” *Id.* (citing *Arnstein v. Porter*, 154 F.2d 464, 469 (2d Cir. 1946)). The Central District again used the *Skidmore* standard when explaining the many ways in which access can be proven “including merely by demonstrating that the work is ‘available on demand.’” *DuMond v. Reilly*, No. CV 19-8922-GW-AGR_x, 2021 U.S. Dist. LEXIS 199398, at *16 (C.D. Cal. Jan. 8, 2021). In 2023, the Central District again stated that “access can be proven through ‘a trivial showing’ of widespread dissemination.” *TeamLab Inc. v. Museum of Dream Space, LLC*, 650 F. Supp. 3d 934, 951 (C.D. Cal. 2023). These

consistent applications of the *Skidmore* standard underscore its enduring significance in establishing access in copyright cases.

Additionally, previous cases in other jurisdictions have discussed how easy a showing of access can be due to the public availability of websites on the internet. The Pennsylvania Western District Court stated that proving access was not a difficult task “due to the ease with which users can view web sites.” *Conf. Archives, Inc. v. Sound Images, Inc.*, No. 3:2006-76, 2010 U.S. Dist. LEXIS 46955, at *32 (W.D. Pa. Mar. 31, 2010). A Florida court did not even bother addressing the access issue “given that anyone with access to the internet had access to [the] website.” *Green Bullion Financial Servs., LLC v. Money4Gold Holdings, Inc.*, 639 F. Supp. 2d 1356, 1361 (S.D. Fla. 2009). Even a court in California declared that the plaintiff had proven access simply “because [plaintiff’s] website was publicly available on the internet.” *Allen v. Ghoulish Gallery*, No. 06cv371 NLS, 2007 U.S. Dist. LEXIS 86224, at *60 (S.D. Cal. Nov. 20, 2007). In light of these precedents, it is evident that the accessibility of online content has significantly influenced the threshold for proving access in copyright infringement cases. Various jurisdictions have acknowledged the ease with which individuals can view content on the internet, emphasizing the challenge in disputing access when content is publicly available.

Scholars have also weighed in on the issue of access in a technologically advanced world. Historically, establishing defendant’s access to a plaintiff’s work

required much more effort and evidence, but, as recent court opinions have recognized, “modern mass media have changed these probabilities.” Christopher Buccafusco, *There’s No Such Thing as Independent Creation, and It’s a Good Thing, Too*, 64 Wm. & Mary L. Rev. 1617 (2023).⁷⁵ The internet and its innumerable platforms “give creators access to huge swaths of the world’s creative production.” *Id.* at 1655. This access “coupled with the passive nature of much online media consumption, where algorithms and playlists feed new content without affirmative requests by the consumer, it is possible that people may be exposed to an enormous variety of content.” *Id.* However, this “increased accessibility and consumption of music poses a particular problem for assessing copyright infringement in music.” Emily Ranger-Murdock, “*Blurred Lines*” to “*Stairway to Heaven*”: *Applicability of Selection and Arrangement Infringement Actions in Musical Compositions*, 67 UCLA L. Rev. 1066 (2020).

Courts recognize that access is often established through “the digitally interconnected world and the various ways to access media using platforms like YouTube, Netflix, and Spotify.” Rebecca Alba, *Closing the Stairway to Heaven: The Ninth Circuit Disincentivizes Music Creation*, 23 Tul. J. Tech. & Intell. Prop. 197 (2021).⁷⁶ Authors have explained how because of society’s dependence on

⁷⁵ https://wmlawreview.org/sites/default/files/wmlr_64-6_Buccafusco-pgs1617-1676.pdf

⁷⁶ <https://journals.tulane.edu/TIP/article/view/3656>

technology, “access is increasingly diluted through the Internet and various streaming sites.” Alba, *supra* at 204. Because of this dependence, the level of access is harder to determine in modern times, qualitatively or quantitatively. SaiPranay Vellala, *Conquering Copyright: Why Copyright Needs to be Modernized Based on Practical Illustrations of Inconsistent Copyright Precedent*, 56 Akron L. Rev. 409 (2022).⁷⁷ Though this is not to suggest “the access prong will never protect defendants from spurious lawsuits, but, over time, it will prove less and less helpful.” Buccafusco, *supra* at 1655. As technology increasingly provides a ubiquity of ‘access’ to media, one scholar suggests that “the courts should retain the concept of access not as a factor in the infringement test, but rather as an affirmative defense.” Christina R. Dimeo, *Rethinking Music Copyright Infringement in the Digital World: Proposing a Streamlined Test After the Demise of the Inverse Ratio Rule*, 55 Univ. Rich. L. Rev. 1077, 1102-03 (2021).⁷⁸

This technological shift in society challenges traditional notions of copyright protection and suggests a need for modernized legal frameworks to address the complexities of it all.

/ / /

⁷⁷

<https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=2572&context=akronlawreview>

⁷⁸ <https://scholarship.richmond.edu/lawreview/vol55/iss3/9>

D. THE USE OF ARTIFICIAL INTELLIGENCE PROGRAMS HAS MASSIVE IMPLICATIONS FOR COPYRIGHT INFRINGEMENT CASES

The issue regarding the use of AI programs upon large data sets (such as Facebook’s user information provided to its partners) is hot button topic. As explained in a Congressional Research Report (“CRS”) dated September 29, 2023,

So-called “generative AI” computer programs—such as Open AI’s DALL-E and ChatGPT programs, Stability AI’s Stable Diffusion program, and Midjourney’s self-titled program—are able to generate new images, texts, and other content (or “outputs”) in response to a user’s textual prompts (or “inputs”). These generative AI programs are trained to generate such outputs partly by exposing them to large quantities of existing works such as writings, photos, paintings, and other artworks.

Congressional Research Services (September 29, 2023) *Generative Artificial Intelligence and Copyright Law* (CRS Report No. LSB10922), at 1 (“CRS Report”)⁷⁹

The foregoing CRS Report notes that there is no well-settled law or judicial interpretation regarding key concepts in AI-generated material such as authorship and whether particular uses of AI programming can result in infringement, while others would be considered ‘transformative’ or a ‘fair use.’ *Id.* at 1-3. Although “courts have uniformly declined to recognize copyright in works created absent any human involvement,” including as created by AI programming (*see Thaler v. Perlmutter*, 2023 U.S. Dist. LEXIS 145823, at *20), for purposes of this Appeal, the only issue is access. To that end, the CRS found that in an infringement

⁷⁹ <https://crsreports.congress.gov/product/pdf/LSB/LSB10922>

analysis, access is readily apparent if the AI program was ‘trained’ or ‘exposed’ to a particular data set, which included a plaintiff’s material.

AI systems are “trained” to create literary, visual, and other artistic works by exposing the program to large amounts of data, which may include text, images, and other works downloaded from the internet. This training process involves making digital copies of existing works. As the U.S. Patent and Trademark Office has described, this process “will almost by definition involve the reproduction of entire works or substantial portions thereof.”
CRS Report, *supra* at 3 (citing USPTO, *Public Views on Artificial Intelligence and Intellectual Property Policy*, at 24 (October 2020))

One artificial intelligence company, Open AI, “acknowledges that its programs are trained on ‘large publicly available datasets that include copyrighted works . . . [c]reating such copies without permission may infringe the copyright holders’ exclusive right to make reproductions of their work.” *Id.* The CRS report further states that “For AI outputs, access might be shown by evidence that the AI program was trained using the underlying work. For instance, the underlying work might be part of a publicly accessible internet site that was downloaded or ‘scraped’ to train the AI program.” *Id.* In the present action, the underlying work, i.e., Appellant’s September 23, 2014 post including the lyrics to *Upgraded 2.0*, may undoubtedly be part of a *privately* accessible data set maintained by Facebook and shared to its partners.

Whether publicly or privately accessible, a key finding by the CRS is that a data set used in the training of an AI program is unquestionably ‘accessed’ for purposes of an infringement analysis. *Id.* As such, a USPTO report found that

there must be a solution regarding the “need to protect the data associated with AI” and further that there are “specifically gaps in IP protection for the trained model and its associated coefficients.” USPTO, *Public Views on Artificial Intelligence and Intellectual Property Policy*, at 15 (October 2020).⁸⁰ Those gaps in protection have already started to be filled by plaintiffs seeking to protect their rights.

Multiple lawsuits are currently pending with plaintiffs “claiming the training process for AI programs infringed their copyrights in written and visual works” (CRS Report, *supra* at 3), including several against Meta Platforms, (formerly known as Facebook) for its use of the plaintiffs’ copyrighted materials in ‘training’ its AI program, known as “Llama,” which was released in February 2023. *Id*; see also Complaint, Dkt. 1, *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417 (N.D. Cal. 2023)) and Complaint, Dkt. 1, *Chabon v. Meta Platforms, Inc.*, 3:23-cv-04663 (N.D. Cal. 2023). The *Kadrey* and *Chabon* matters filed in the Northern District of California were ordered consolidated on December 7, 2023, and notably, the plaintiffs’ claim for direct copyright infringement survived a 12(b)(6) motion to dismiss.

E. THE CHAIN OF EVENTS AS DESCRIBED HEREINABOVE SHOWS A REASONABLE POSSIBILITY OF ACCESS

In its February 23, 2022 Order, dismissing the FAC, the District Court found

⁸⁰ https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf

that plaintiff may prove access by establishing “a chain of events linking the plaintiff’s work and the defendant’s access.” (ER_024) *citing* Art Attacks, 581 F.3d at 1143. “At the pleading stage, a plaintiff must plead facts that, if accepted as true, “show a reasonable possibility, not merely a bare possibility, that an alleged infringer had the chance to view the protected work.” *Id.* The District Court reviewed the FAC allegations regarding the Facebook/UMG chain of events and found that “Plaintiff fails to allege facts showing each link in the chain of events leading to *Moving Defendants’* access.” (ER_026), and thus, the allegations of access based on UMG’s alleged agreement with Facebook is speculative, conclusory, and thus insufficient to plausibly plead access”

To briefly reiterate, the *Kadrey* and *Chabon* matters pending in the Northern District of California survived a 12(b)(6) motion to dismiss with respect to their direct copyright infringement claim against Meta Platforms. *See generally* Order Granting Motion to Dismiss, Dkt. 56, *Kadrey supra*, 3:23-cv-03417. The difference between Appellant and the aforementioned plaintiffs is merely the shroud under which Appellant’s data may have been used, and the degrees of separation between Appellees’ alleged use of the data, and the Appellant’s posting of it. In *Kadrey*, Meta’s publicly available information regarding how the Llama AI program was trained clearly admitted that it used data sets, which included plaintiffs’ copyrighted works. *See* First Consolidated Amended Complaint, Dkt. 64, at ¶¶ 31-39. *Kadrey, supra*, 3:23-cv-03417. The action is between the owners

of those copyrighted works and the company that owns the AI program – a clear one to one connection.

Here, however, Appellant’s allegations are that her Work was included in Facebook’s repository of data, which was improperly shared in a surreptitious manner with its partners, including Appellee UMG (ER_077-78, ER_087 ¶ 44 – ER_094 ¶ 56). This type of data sharing was intentionally concealed from the public by Facebook (or possibly negligently or recklessly permitted), as Facebook was barred by the FTC since at least the 2012 from misrepresenting its privacy controls *See* Stip. Ord. for Civ. Penalty, Monetary Jdgmt., and Inj. Relief, Dkt. 2-1, at 3-4, *Facebook, supra* 19-cv-2184. Thereafter, UMG likely *privately* used search queries, data scraping tools or AI programs on the Facebook data it was not supposed to have possession over (*see supra*, Section IV(B)(3)-(4)). Again, another possibly surreptitious act, which would have been concealed from the public; nevertheless, the technological landscape as described hereinabove does not make the chain of events any less likely. As such, the Court should reconsider the reasonability of the chain of events pleaded, reverse and remand the action to the District Court, and permit Appellants to engage in discovery on these issues.

F. THE DISTRICT COURT ABUSED ITS DISCRETION IN FAILING TO GRANT APPELLANTS LEAVE TO AMEND THE COMPLAINT

At the pleading stage, dismissal with prejudice is appropriate only if a complaint “could not be saved by any amendment.” *Astor-White*, 733 F. App’x at

407 (9th Cir. 2018) (quoting *Leadsinger, Inc. v. BMG Rights Publ'g*, 512 F.3d 522, 532 (9th Cir. 2008). “An outright refusal to grant leave to amend without a justifying reason is . . . an abuse of discretion.” *Manzarek v. St. Paul Fire & Marine, Ins. Co.*, 519 F.3d 1025, 1034 (9th Cir. 2008).

Here, failure to grant leave to amend was an abuse of discretion because (1) the reasonability of Appellant’s theory of access is a factual determination, which could have been buttressed by further allegations and (2) Appellants were deprived of the opportunity to conduct discovery, including but not limited to (a) whether Facebook, in fact, made any of its user data accessible to Appellees; and (b) whether Appellees accessed or utilized search tools or AI programs in connection with the Facebook user information. Finally, the offer of proof with respect to both a Second Amended Complaint and what discovery is likely to reveal, as set forth above, makes the failure to grant leave an abuse of discretion. Accordingly, the District Court abused its discretion in denying Appellants’ leave to amend.

VII. CONCLUSION

The trial court was incorrect in granting Appellees’ 12(b)(6) motions to dismiss. For the foregoing reasons, the judgment of the district court should be reversed, or remanded for further consideration consistent with the correct law as set forth herein.

Date: April 23, 2024

Steven T. Lowe, Esq.

/s/ Steven T. Lowe, Esq.

Steven T. Lowe, Esq.

Attorney for Appellant Michele Ronk

STATEMENT OF RELATED CASES

I do not know of any other related cases in this Court.

Date: April 23, 2024

Steven T. Lowe, Esq.

/s/ Steven T. Lowe, Esq.

Steven T. Lowe, Esq.

Attorney for Appellants Michele Ronk

CERTIFICATE OF COMPLIANCE

Pursuant to Fed. R. App. P. 32(a)(7)(C), I certify that:

This brief complies with the type-volume limitation of Fed. R. App. P. 32(a)(7)(B) because this brief contains 11,759 words, excluding the parts of the brief exempted by Fed. R. App. P. 32(a)(7)(B)(iii).

This brief complies with the typeface requirements of Fed. R. App. P. 32(a)(5) and the type style requirements of Fed. R. App. P. 32(a)(6) because this brief has been prepared in a proportionately spaced typeface using Microsoft Word version 15.24 and Times New Roman 14-point font.

Date: April 23, 2024

Steven T. Lowe, Esq.

/s/ Steven T. Lowe, Esq.
Steven T. Lowe, Esq.

Attorney for Appellant Michele Ronk

CERTIFICATE OF SERVICE

I hereby certify that on April 23, 2024 I electronically filed the foregoing with the Clerk of the Court for the United States Court of Appeals for the Ninth Circuit by using the appellate CM/ECF system.

Participants in the case who are registered CM/ECF users will be served by the appellate CM/ECF system.

Date: April 23, 2024

Steven T. Lowe, Esq.

/s/ Steven T. Lowe, Esq.
Steven T. Lowe, Esq.

Attorney for Appellant Michele Ronk