Washington Law Review

Volume 93 | Number 2

6-1-2018

How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem

Amanda Levendowski

HOW COPYRIGHT LAW CAN FIX ARTIFICIAL INTELLIGENCE'S IMPLICIT BIAS PROBLEM

Amanda Levendowski*

Abstract: As the use of artificial intelligence (AI) continues to spread, we have seen an increase in examples of AI systems reflecting or exacerbating societal bias, from racist facial recognition to sexist natural language processing. These biases threaten to overshadow AI's technological gains and potential benefits. While legal and computer science scholars have analyzed many sources of bias, including the unexamined assumptions of its often-homogenous creators, flawed algorithms, and incomplete datasets, the role of the law itself has been largely ignored. Yet just as code and culture play significant roles in how AI agents learn about and act in the world, so too do the laws that govern them. This Article is the first to examine perhaps the most powerful law impacting AI bias: copyright.

Artificial intelligence often learns to "think" by reading, viewing, and listening to copies of human works. This Article first explores the problem of bias through the lens of copyright doctrine, looking at how the law's exclusion of access to certain copyrighted source materials may create or promote biased AI systems. Copyright law limits bias mitigation techniques, such as testing AI through reverse engineering, algorithmic accountability processes, and competing to convert customers. The rules of copyright law also privilege access to certain works over others, encouraging AI creators to use easily available, legally low-risk sources of data for teaching AI, even when those data are demonstrably biased. Second, it examines how a different part of copyright law—the fair use doctrine—has traditionally been used to address similar concerns in other technological fields, and asks whether it is equally capable of addressing them in the field of AI bias. The Article ultimately concludes that it is, in large part because the normative values embedded within traditional fair use ultimately align with the goals of mitigating AI bias and, quite literally, creating fairer AI systems.

---

* Technology Law and Policy Clinical Teaching Fellow, New York University School of Law and Research Fellow, NYU Information Law Institute. Thanks to Solon Barocas, Barton Beebe, Ilya Beylin, Kiel Brennan-Marquez, Miles Brundage, Ryan Calo, Rochelle Dreyfuss, Seth Endo, Jeanne Fromer, Gautam Hans, Anne Hassett, H. Brian Holland, Sonia Katyal, Craig Konnoth, Karen Levy, Tiffany Li, Richard Loren Jolly, Laurel Reik, Zvi Rosen, Jason Schultz, Scott Skinner-Thompson, Naomi Sunshine, Chris Sprigman, Eva Subotnik, Ari Waldman, Clinton Wallace, Rebecca Wexler, Felix Wu, and Cameron Tepski for their thoughtful observations and suggestions. Rachel Brooke provided stellar research assistance. Thanks also to the editors of the *Washington Law Review* for their insightful recommendations. I am also grateful to the organizers and participants of the 2017 Works-in-Progress in Intellectual Property Colloquium, the NYU Lawyering Scholarship Colloquium, We Robot 2017, the NYU Law Engelberg Center Faculty Workshop, the Intellectual Property Scholars Conference, and the Shefelman Faculty Workshop at the University of Washington, where I presented drafts of this paper. The NYU Technology Law and Policy Clinic filed amicus briefs in several cases discussed in this Article, and I previously worked for firms involved with other cases. Any discussion is based on public information and all views expressed are my own.

INTRODUCTION

In 2013, Google announced the release of word2vec, a toolkit capable of representing how words are used in relation to one another so as to better understand their meanings.[1] Word2vec can recognize that Beijing is to China in the same way as Warsaw is to Poland, as capital and country, but not in the same way as Paris relates to Germany.[2] This

1. Specifically, word2vec "uses distributed representations of text to capture similarities among concepts." Thomas Mikolov et al., *Learning the Meaning Behind Words*, GOOGLE OPEN SOURCE BLOG (Aug. 14, 2013), https://opensource.googleblog.com/2013/08/learning-meaning-behind-words.html [https://perma.cc/AX9C-ME4X] [hereinafter *Learning the Meaning Behind Words*].

2. *Id.*

technique, called "word embedding,"[3] plays a role in many downstream uses of artificial intelligence (AI) tasks; Google uses it to improve its search engine, image recognition, and email auto-response tools.[4] Since its launch, word2vec has become one of the most popular embedding models.[5]

There is a significant problem with word2vec: it is sexist. More specifically, word2vec reflects the gendered bias embedded in the Google News corpus used to train it.[6] In 2016, researchers from Boston University and Microsoft Research New England uncovered that word2vec was riddled with gender bias exemplified by a particularly noteworthy word embedding, which projected that man is to computer programmer in the same way that woman is to homemaker.[7] Word embeddings are used in many downstream AI tasks, including improving web search. Thus, if an underlying dataset reflects gendered bias, those biases would be reinforced and amplified by sexist search results that, for example, rank results for computer programmers with male-sounding

---

3. *See* Yoshua Bengio et al., *A Neural Probabilistic Language Model*, 3 J. MACH. LEARNING RES. 1137–55 (2003), http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf [https://perma.cc/LC2C-QYUK].

4. Jeff Larson et al., *Breaking the Black Box: How Machines Learn to Be Racist*, PROPUBLICA (Oct. 19, 2016), https://www.propublica.org/article/breaking-the-black box-how-machines-learn-to-be-racist?word=blackness [https://perma.cc/D2S4-L2NJ]. Word2vec has also been used to improve other search queries and parse curriculum vitae. *See* Dwaipayan Roy et al., *Using Word Embeddings for Automatic Query Expansion*, 16 SIGIR WORKSHOP ON NEURAL INFO. RET. (July 21, 2016), https://arxiv.org/abs/1606.07608 [https://perma.cc/8KC3-BFUD] (search queries); Melanie Tosik et al., *Word Embeddings vs Word Types for Sequence Labeling: The Curious Case of CV Parsing*, PROCEEDINGS OF NAACL-HLT (2015), http://www.aclweb.org/anthology/W15-1517 [https://perma.cc/9NSC-53QR] (curriculum vitae). Both examples are mentioned in TOLGA BOLUKBASI ET AL., arXiv:1607.05620, MAN IS TO COMPUTER PROGRAMMER AS WOMAN IS TO HOMEMAKER? DEBIASING WORD EMBEDDINGS (2016), https://arxiv.org/abs/1607.06520 [https://perma.cc/45UY-WJD8] [hereinafter *Debiasing Word Embeddings*].

5. *See* Emerging Technology from the arXiv, *How Vector Space Mathematics Reveals the Hidden Sexism in Language*, MIT TECH. REV. (July 27, 2016), https://www.technologyreview.com/s/602025/how-vector-space-mathematics-reveals-the-hidden-sexism-in-language/ [https://perma.cc/R78V-2TMT].

6. As Cathy O'Neil has proclaimed, "I'll stop calling algorithms racist when you stop anthropomorphizing AI." Cathy O'Neil, *I'll Stop Calling Algorithms Racist when You Stop Anthropomorphizing AI*, MATHBABE (Apr. 7, 2016), https://mathbabe.org/2016/04/07/ill-stop-calling-algorithms-racist-when-you-stop-anthropomorphizing-ai/ [https://perma.cc/9LA6-BTLL].

7. Co-authors Tolga Bolukbasi and Venkatesh Saligrama are affiliated with Boston University; Kai-Wei Chang, James Zou, and Adam Kalai are affiliated with Microsoft Researcher New England. *Debiasing Word Embeddings*, *supra* note 4; *see also* TOLGA BOLUKBASI ET AL., QUANTIFYING AND REDUCING STEREOTYPES IN WORD EMBEDDINGS, arXiv:1606.06121, IMCL WORKSHOP ON #DATA4GOOD: MACH. LEARNING IN SOCIAL GOOD APPLICATIONS 41 (2016), https://arxiv.org/abs/1606.06121 [https://perma.cc/W29Q-N9LT] [hereinafter *Quantifying and Reducing Stereotypes*].

names more highly than those of female-sounding names.[8] "Due to their wide-spread usage as basic features," the researchers warned, "word embeddings not only reflect such stereotypes but can amplify them."[9]

AI systems are commonly "taught" by reading, viewing, and listening to copies of works created by humans. Many of those works are protectable by copyright law.[10] Google, for example, negotiated with multiple global news agencies to license articles for Google News after the company was sued for copyright infringement.[11] For Google, the articles used to create the Google News corpus, which were ultimately used to create word2vec, were easily available and legally low-risk.[12]

Although Google released the word2vec toolkit as open source, the underlying Google News corpus was not released at all.[13] It is all but unimaginable that a researcher could hope to strike comparable licensing deals, even in a bid to create a less biased corpus. And without access to the underlying corpus, downstream researchers cannot examine whether a news outlet or journalist exhibits gender bias across multiple articles, nor could researchers supplement the corpus with data derived from

---

8. *See Debiasing Word Embeddings*, *supra* note 4, at 3. The same can also be true of racial bias. *See* SU LIN BLODGETT & BRENDAN O'CONNOR, arXiv:1707.00060006, RACIAL DISPARITY IN NATURAL LANGUAGE PROCESSING: A CASE STUDY OF SOCIAL MEDIA (2017), https://arxiv.org/abs/1707.00061 [https://perma.cc/WR23-UKEK].

9. *See Debiasing Word Embeddings*, *supra* note 4, at 3. For an in-depth exploration of how easily implicit biases encoded in human language are picked up by AI, see Aylin Caliskan, et al., *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases*, 356 SCI. 183 (2017), http://science.sciencemag.org/content/356/6334/183 [https://perma.cc/FXS3-SFAP]. And for a related exploration of how implicit biases encoded in images can be amplified by AI systems, see Jieyu Zhao et al., *Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints*, https://homes.cs.washington.edu/~my89/publications/bias.pdf [https://perma.cc/HA9U-KQT3]; Tom Simonite, *Machines Taught by Photos Learn a Sexist View of Women*, WIRED (Aug. 21, 2017), https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women [https://perma.cc/L8KC-CK5R].

10. *See generally* 17 U.S.C. § 106 (2012).

11. Agence France Presse v. Google Inc., No.1:05CV00546, 2005 WL 5834897 (D.D.C. Apr. 29, 2005). The parties settled for an undisclosed amount; Google continued to expand its licensing agreements with other content providers. Josh Cohen, *More Hosted News Partners in Europe*, GOOGLE NEWS BLOG (Mar. 17, 2009), https://news.googleblog.com/2009/03/more-hosted-news-partners-in-europe.html [https://perma.cc/N8E3-ULQC]; Josh Cohen, *Original Stories, from the Source*, GOOGLE NEWS BLOG (Aug. 31, 2007), https://news.googleblog.com/2007/08/original-stories-from-source.html [https://perma.cc/4TNR-KCX6] (announcing feature to send readers "directly to the original source" of a news story); Josh Cohen, *Extending the Associated Press as Hosted News Partner*, GOOGLE NEWS BLOG (Aug. 30, 2010), https://news.googleblog.com/2010/08/extending-associated-press-as-hosted.html [https://perma.cc/TBP6-CUGE].

12. MIKOLOV ET AL., arXiv:1301.3781v3, EFFICIENT ESTIMATION OF WORD REPRESENTATIONS IN VECTOR SPACE (2013), https://arxiv.org/abs/1301.3781 [https://perma.cc/5XFZ-8D9P].

13. *word2vec*, GOOGLE CODE ARCHIVE (July 29, 2013), https://code.google.com/archive/p/word2vec/ [https://perma.cc/EZF6-KW63].

additional, less biased works. Indeed, as the researchers who identified the biases embedded in the Google News corpus noted, locking up the dataset makes it "impracticable and even impossible . . . to reduce the [biased] stereotypes during the training of the word vectors."[14]

Even as our banks and our bosses,[15] our cars and our courts[16] increasingly adopt AI, bias remains a significant and complex problem.[17] One source of bias in AI systems is, as exemplified by word2vec, data that reflect implicit bias. Indeed, as the Obama White House aptly identified in its whitepaper on AI, "AI needs good data. If the data is incomplete or biased, AI can exacerbate problems of bias."[18] AI's largely homogenous community of creators, which skews toward white men, is another source of bias.[19] Flawed algorithms can also contribute to bias, evident in Google search algorithms that featured Barbie as the

14. *Quantifying and Reducing Stereotypes*, *supra* note 7, at 43.

15. Charles Lane, *Will Using Artificial Intelligence to Make Loans Trade One Kind of Bias for Another?*, NPR: MORNING EDITION (Mar. 31, 2017, 5:06 AM), http://www.npr.org/sections/alltechconsidered/2017/03/31/521946210/will-using-artificial-intelligence-to-make-loans-trade-one-kind-of-bias-for-anot [https://perma.cc/R9MT-ELJG]; Ted Greenwald, *How AI Is Transforming the Workplace*, WALL ST. J. (Mar. 10, 2017), https://www.wsj.com/articles/how-ai-is-transforming-the-workplace-1489371060 [https://perma.cc/7YTA-GMP5].

16. PETER STONE & RODNEY BROOKES, ET AL., STANFORD, "ARTIFICIAL INTELLIGENCE AND LIFE IN 2030," ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE: REPORT OF THE 2015–2016 STUDY PANEL, 18–23 (2016), https://ai100.stanford.edu/sites/default/files/ai100report10032016fnl_singles.pdf [https://perma.cc/97UL-JYHW]; Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/SSQ7-Y8TW].

17. For a comprehensive taxonomy of technical sources of bias in data mining and AI, see Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 671–73 (2016).

18. EXEC. OFFICE OF THE PRESIDENT, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf [https://perma.cc/YN5X-KKAX].

19. Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016), https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html [https://perma.cc/AN63-XYS6]. In 2016, it was estimated that fewer than 14% of the workforce in machine learning, the largest commercial subfield of AI, were women. Cale Guthrie Weissmam, *The Women Changing the Face of AI*, FAST CO. (Aug. 18, 2016), https://www.fastcompany.com/3062932/mind-and-machine/ai-is-a-male-dominated-field-but-an-important-group-of-women-is-changing-th [https://perma.cc/V62V-HSLS]. This gender imbalance is reinforced by job listings in AI: a recent analysis of 78,768 engineering job listings found that postings for software engineers in the machine intelligence category had a gender-bias score favoring men more than twice as high as the next category. Kieran Snyder, *Language in Your Job Post Predicts the Gender of Your Hire*, TEXTIO (June 21, 2016), https://textio.ai/gendered-language-in-your-job-post-predicts-the-gender-of-the-personyoull-hire-cd150452407d#.rht0s16ov [https://perma.cc/9CWC-RA49] ("In light of the bias distributions above, the apparent scarcity of women in machine intelligence jobs is probably more than anecdotal.").

lone woman in top image results for CEO,[20] or serve up ads implying the existence of criminal records when running for black-sounding names.[21] Incomplete datasets are another common source of bias, particularly datasets that fail to reflect a diversity of facial features and skin tones.[22]

---

20. Notably, the Barbie featured in the search results was not even a real Mattel product—she was created by satirical outlet *The Onion* for a story, which dryly observed that "women don't run companies." T.C. Sottek, *Google Search Thinks the Most Important Female CEO Is Barbie*, VERGE (Apr. 9, 2015), https://www.theverge.com/tldr/2015/4/9/8378745/i-see-white-people [https://perma.cc/S42G-EQAF]. The biases reflected and reinforced through image search results have measurable effects on people's perception. *See* MATTHEW KAY ET AL., UNEQUAL REPRESENTATION AND GENDER STEREOTYPES IN IMAGE SEARCH RESULTS FOR OCCUPATIONS (2015), https://dub.washington.edu/djangosite/media/papers/unequalrepresentation.pdf [https://perma.cc/5DJT-WUWY]; Megan Rose Dickey, *Algorithmic Accountability*, TECH CRUNCH (Apr. 30, 2017), https://techcrunch.com/2017/04/30/algorithmic-accountability/ [https://perma.cc/Y5AU-8MRZ] (discussing analysis of former Googler Sorelle Friedler's perspective on the philosophy and consequences of whitewashed Google Images search results for "person").

21. Latanya Sweeney, *Discrimination in Online Ad Delivery*, 56 COMM. OF THE ACM 5, 44–54 (2013), https://dataprivacylab.org/projects/onlineads/1071-1.pdf [https://perma.cc/X7NR-B8JZ]. Researchers from Carnegie Mellon also uncovered gender bias is Google's ads, which churned out fewer ads for high-paying jobs to women than it did to men. AMIT DATTA ET AL., AUTOMATED EXPERIMENTS ON AD PRIVACY SETTINGS, PROCEEDINGS ON PRIVACY ENHANCING TECH. 92, 93 (2015), https://www.degruyter.com/downloadpdf/j/popets.2015.1.issue-1/popets-2015-0007/popets-2015-0007.pdf [https://perma.cc/PHQ8-M7YP].

22. *See* Joz Wang (@jozjozjoz), FLICKR (May 10, 2009), https://www.flickr.com/photos/jozjozjoz/3529106844 [https://perma.cc/E92F-EAZQ] (Nikon camera identified a Taiwanese-American blogger as "blinking"); Jacky Alciné (@jackyalcine), TWITTER (June 28, 2015, 6:22 PM), https://twitter.com/jackyalcine/status/615329515909156865?lang=en [https://perma.cc/4KHM-2PM5] ("Google Photos, y'all fucked up. My friend's not a gorilla," wrote Alciné in response to Google Photos tagging two of his black friends as "gorillas"); Tom Simonite, *When It Comes to Gorillas Google Photos Remains Blind*, WIRED (Jan. 11, 2018), https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/ [https://perma.cc/C5L6-JP9G] (noting that Google's response to the error was to remove the "gorilla" tag entirely). Researchers have found that implicit biases of facial recognition are affected by where AI systems are designed and who designs them. *See* P. Jonathan Phillips et al., *An Other-Race Effect for Facial Recognition Algorithms*, 8 TAP 14 (2011), https://pdfs.semanticscholar.org/5fb9/4eaf71196ed9cb87bbd881af34be8b4bc919.pdf [https://perma.cc/8U3D-LSN8] (reporting that facial algorithms from Western countries recognized Caucasian faces more accurately than East Asian faces, and that East Asian algorithms similarly recognized East Asian faces more accurately than Caucasian ones).

Government AI systems, which have an admittedly different data-collection calculus, have also been demonstrated to show bias. In a particularly troubling example, researchers—including a technologist for the Federal Bureau of Investigation—analyzed multiple types of facial recognition algorithms and concluded that the AI had biased matching accuracies, systemically struggling to identify black, female, and young faces. *See* Brendan F. Klare & Mark J. Burge, *Face Recognition Performance: Role of Demographic Information*, 7 IEEE TRANS. ON INFO. FORENSICS & SECURITY 6 (2012), http://ieeexplore.ieee.org/document/6327355/authors?ctx=authors [https://perma.cc/RTF6-9KEV]. In January 2017, President Donald Trump signed an Executive Order that called on travelers to be screened to evaluate the likelihood of each person becoming "a positively contributing member of society" who would "make contributions to the national interest." Exec. Order No. 13769, 82 Fed. Reg. 8977 (Jan. 27, 2017). Recently, Immigration and Customs Enforcement (ICE) proposed to automate extreme vetting by relying on AI systems to scan the

Commercial facial detection AI systems, for example, have been plagued with racial bias.[23] In 2017, a mobile app called FaceApp introduced a "hot" photo editing feature that conflated attractiveness with whiteness by automatically lightening users' skin tones in photos, which the CEO attributed to "an unfortunate side-effect of the underlying neural network caused by training set bias."[24]

Even an AI system designed by diverse creators and driven by impeccable algorithms will nevertheless generate biased results if reliant on biased data. Computer scientists tend to put this axiom more bluntly: garbage in, garbage out.[25]

---

internet, such as social media profiles, to predict which individuals are likely to satisfy the aforementioned clause and, based on those predictions, identify individuals for deportation or visa denial. For a critique of the difficulties in using AI systems to accurately identify elusive concepts like productivity to society, including dataset-related challenges, see the Brennan Center for Justice's extreme vetting resources. *ICE Extreme Vetting Initiative: A Resource Page*, BRENNAN CTR. FOR JUSTICE N.Y.U., https://www.brennancenter.org/analysis/ice-extreme-vetting-initiative-resource-page [https://perma.cc/HW39-L8TS].

Bias in government facial recognition AI can have outsized impacts, including misidentifying a suspect or mistakenly identifying an innocent person. Clare Garvie & Alvaro Bedoya et al., *The Perpetual Line-Up: Unregulated Facial Recognition in America*, GEO. CTR ON PRIVACY & TECH. (Oct. 18, 2016), https://www.perpetuallineup.org/ [https://perma.cc/8TBK-WS8K]. The disproportionate negative impact of facial recognition surveillance on people of color is significant, and AI's role in automating the status quo can reflect the ways in which American surveillance is rooted in anti-blackness. For a stunning historical accounting of surveillance and anti-blackness, see SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS (2016).

23. Joy Buolamwini and Timnit Gebru recently created the Pilot Parliaments Benchmark dataset to test the accuracy of three commercial gender classification AI systems—by IBM, Microsoft, and Face++—and empirically demonstrated the disproportionately high error rates for darker-skinned females in the dataset. *See* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROCEEDINGS OF MACH. LEARNING RES. 1–15 (2018), CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY (2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf [https://perma.cc/3CGB-ZX8S]. Notably, copyright law played a significant role in determining the images used in the intersectional benchmark dataset: Buolamwini and Gebru used images of global parliamentarians because "they are public figures with known identities and photos available under non-restrictive licenses [are] posted on government websites." *Id.* at 5.

24. Elena Cresci, *FaceApp Apologises for 'Racist' Filter that Lightens Users' Skintone*, GUARDIAN (Apr. 25, 2017), https://www.theguardian.com/technology/2017/apr/25/faceapp-apologises-for-racist-filter-which-lightens-users-skintone [https://perma.cc/5LQT-DYYH]; *see also* Selina Cheng, *An Algorithm Rejected an Asian Man's Passport Photo for Having "Closed Eyes,"* QUARTZ (Dec. 7, 2016), https://qz.com/857122/an-algorithm-rejected-an-asian-mans-passport-photo-for-having-closed-eyes/ [https://perma.cc/VYE4-2Q6V] (Taiwanese engineering student unable to renew his New Zealand passport online because his eyes were identified as "closed").

25. The garbage in, garbage out quandary is as old as the first computer. Charles Babbage, the inventor and philosopher credited with creating the first mechanical computer, first addressed the issue back in 1864:

> On two occasions I have been asked—'Pray Mr. Babbage, if you put into the machine the wrong figures, . . . will the right answers come out?' . . . . I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

Legal and computer science scholars—including Batya Friedman, Helen Nissenbaum, Latanya Sweeney, Danielle Citron, Frank Pasquale, Paul Ohm, Kate Crawford, Solon Barocas, and Andrew Selbst, among many others—have long examined sources of computational bias,[26] as well as the ethical dilemmas posed by collecting, storing, and using the vast quantities of "Big Data" necessary to train most AI.[27] Indeed, there is a robust body of scholarship,[28] even entire conferences and research

---

CHARLES BABBAGE, PASSAGES FROM THE LIFE OF A PHILOSOPHER 67 (1864), https://archive.org/stream/passagesfromlif01babbgoog/passagesfromlif01babbgoog_djvu.txt [https://perma.cc/3LV9-2TRX]. I would be remiss to mention Babbage without acknowledging the contributions of Ada Lovelace, who wrote the first algorithm meant to be executed by a machine and is often identified as the first computer programmer. For an introduction to Ada Lovelace's contributions, see Eugene Eric Kim & Betty Alexandra Toole, *Ada and the First Computer*, 280 SCI. AM. 76 (1999), http://www.cs.virginia.edu/~robins/Ada_and_the_First_Computer.pdf [https://perma.cc/2TNU-9KPN]. For a longer examination of Lovelace's role in kickstarting the contemporary computer age, see BETTY A. TOOLE, ADA: THE ENCHANTRESS OF NUMBERS, PROPHET OF THE COMPUTER AGE (1998).

26. *See* Batya Friedman & Helen Nissenbaum, *Bias and Computer Systems*, 14 ACM TRANSACTIONS ON INFO. SYS. 3, 330 (1996), http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=EC7E61E26D3061457B91E82D14066EC3?doi=10.1.1.93.9237&rep=rep1&type=pdf [https://perma.cc/ETS9-NREQ]; Sweeney, *supra* note 21; Danielle Keats Citron, *Technological Due Process*, 85 WASH. L. REV. 1249 (2008); Danielle Keats Citron & Frank A. Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); Paul Ohm, *The Underwhelming Benefits of Big Data*, 161 U. PA. L. REV. ONLINE 339, 340 (2013), https://www.pennlawreview.com/online/161-U-Pa-L-Rev-Online-339.pdf [https://perma.cc/U3FS-B9M8] (enumerating the "bad outcomes" that will inevitably follow from reliance and overreliance on big data); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 119–20 (2014) (exploring how personal privacy harms, such as revealing sensitive personal information, may stem from use of one's data without knowledge or consent); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 14 CALIF. L. REV. 671, 671–73 (2016) (computer scientist and legal scholar offering a thorough taxonomy of technical sources of bias in data mining).

27. *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); Neil M. Richards & Johnathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 41–42 (2013) (delving into ways that data collectors are empowered at the expensive of individual privacy and identity, even as collection methodologies and analyses are "shrouded in legal and commercial secrecy"); Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014); Faisal Kamiran & Toon Calders, *Data Preprocessing Techniques for Classification Without Discrimination*, 33 KNOWLEDGE & INFO. SYS. 1, 1 (2012), https://link.springer.com/article/10.1007/s10115-011-0463-8 [https://perma.cc/S32W-N8DR] (suggesting algorithmic solutions—specifically suppression, massaging, and reweighing or resampling of sensitive attributes or discriminatory proxies—for reprocessing data to avoid biased classifications).

28. *See* NINA GRIĆ-HLAČA ET AL., THE CASE FOR PROCESS FAIRNESS IN LEARNING: FEATURE SELECTION FOR FAIR DECISION MAKING, CONF. ON NEURAL INFO. PROCESSING SYS. (2016), http://www.mlandthelaw.org/papers/grgic.pdf [https://perma.cc/5NRZ-LTQT]; Ed Felten & Terah Lyons, *Public Input and Next Steps on the Future of Artificial Intelligence*, MEDIUM (Sept. 6, 2016), https://medium.com/@USCTO/public-inputand-next-steps-on-the-future-of-artificialintelligence-458b82059fc3 [https://perma.cc/2YJJ-CE8H]; Kate Crawford & Ryan Calo, *There is a Blind Spot in AI Research*, 538 NATURE 311 (2016).

institutes,[29] dedicated to reducing bias in AI. And there is a growing body of work suggesting that our focus ought to be on limiting or eliminating AI systems that may be used against marginalized communities rather than making those systems less biased.[30] Largely absent from the scholarship is a discussion of the role that law plays in determining how AI systems are developed and who is empowered to develop them.

Both the Computer Fraud and Abuse Act (CFAA) and trade secret law affect AI systems. Some courts have interpreted the CFAA as prohibiting violation of an employer's computer-use policies or a website's Terms of Service,[31] which can chill algorithmic accountability

29. Since 2014, the FATML and FAT* workshops—FAT being short for Fairness, Accountability, and Transparency—have brought researchers and scholars into conversation around these questions in AI. *See* CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY IN MACHINE LEARNING (2016), http://www.fatml.org/ [https://perma.cc/NF8H-W8QZ]; FAT* (2018), https://fatconference.org/ [https://perma.cc/W79L-K3BC]. The recently launched AI Now Institute, co-founded by Kate Crawford and Meredith Whittaker and housed at New York University, highlights "bias and inclusion" as one of its four core areas of research. *See* KATE CRAWFORD & MEREDITH WHITTAKER, THE AI NOW REPORT: THE SOCIAL AND ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE NEAR TERM (2017).

30. *See, e.g.*, Letter to Axon AI Ethics Board Regarding Ethical Product Development and Law Enforcement (Apr. 26, 2018), https://civilrights.org/axon-product-development-law-enforcement/ [https://perma.cc/B5QT-6YR2] (advocating that Axon, a developer of AI systems targeted to law enforcement, "has a responsibility to ensure that its present and future products, including AI-based products, don't drive unfair or unethical outcomes or amplify racial inequities in policing"); Nabil Hassein, *Against Black Inclusion in Facial Recognition*, DECOLONIZED TECH (Aug. 15, 2017), https://decolonizedtech.com/2017/08/15/against-black-inclusion-in-facial-recognition/ [https://perma.cc/3H66-JBZ9] (discussing the dangerous consequences of "data colonization," or using works created by marginalized persons to diversify biased datasets for training AI systems); Nabil Hassein (@NabilHassein), TWITTER (Aug. 15, 2017, 7:50 PM), https://twitter.com/NabilHassein/status/897651737296855040 [https://perma.cc/3ZES-RQKD] ("But considering who mostly controls facial recognition software, I argue Black folks won't benefit from it getting better at recognizing us."); Simone Browne (@wewatchwatchers), TWITTER (Feb. 10, 2018, 11:36 AM) https://twitter.com/wewatchwatchers/status/962364556218597376 [https://perma.cc/3UEX-K6BP] ("We need a politics of refusal around facial recognition technology (its use to reinforce borders, to criminalize, in capturing 'insurgents,' face reading drones), but that ship has almost completely sailed.").

31. 18 U.S.C. § 1030(a)(2)(C) (2012). The circuit split between broad and narrow interpretations of the CFAA is deep and contentious. *Compare* Brown Jordan Int'l, Inc. v. Carmicle, 846 F.3d 1167, 1174–75 (11th Cir. 2017), *and* United States v. John, 597 F.3d 263, 272 (5th Cir. 2010), *and* Int'l Airport Ctrs., L.L.C. v. Citrin, 440 F.3d 418, 420–21 (7th Cir. 2006), *and* EF Cultural Travel BV v. Explorica, Inc., 274 F.3d 577, 583–84 (1st Cir. 2001) (adopting a broad interpretation of "exceed[ing] authorized access"), *with* United States v. Valle, 807 F.3d 508, 528 (2d Cir. 2015), *and* United States v. Nosal, 676 F.3d 854, 862–63 (9th Cir. 2012), *and* WEC Carolina Energy Sols. LLC v. Miller, 687 F.3d 199, 207 (4th Cir. 2012) (rejecting a broader interpretation). Circuits adopting a narrow interpretation of the CFAA are conscientious that the CFAA creates criminal penalties using the same language as used in the civil provisions, 18 U.S.C. § 1030(g), and have criticized broad interpretations—so much so that broad-interpretation circuits have begun to explicitly address that criticism. Just this summer, for example, the Eleventh Circuit acknowledged that the broad

testing, including digital auditing used to uncover racial discrimination.[32] Indeed, the American Civil Liberties Union is currently litigating on behalf of journalists who have been deterred from such testing, alleging that the CFAA is unconstitutionally overbroad.[33] Trade secret laws, as Rebecca Wexler has meticulously detailed, can shield the code used in algorithms that inform bail, sentencing, and parole decisions from public disclosure.[34] Perhaps the most powerful law channeling the development of AI systems, however, is neither of these: it is copyright law.

---

approach it adopted nearly a decade ago has been widely critiqued by other circuits. EarthCam, Inc. v. OxBlue Corp., No. 15-11893, 2017 WL 3188453, at *9 n.2 (11th Cir. July 27, 2017) ("We decided *Rodriguez* [628 F.3d 1258] in 2010 without the benefit of a national discourse on the CFAA. Since then, several of our sister circuits have roundly criticized decisions like *Rodriguez* because, in their view, simply defining 'authorized access' according to the terms of use of a software or program risks criminalizing everyday behavior . . . . Neither the text, nor the purpose, nor the legislative history of the CFAA, those courts maintain, requires such a draconian outcome. We are, of course, bound by *Rodriguez*, but note its lack of acceptance."). And despite its holding in *Nosal* rejecting a broad interpretation of the CFAA, the Ninth Circuit recently held that continuing to access a website after receiving a cease and desist letter created liability under the CFAA. Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058 (9th Cir. 2016) ("But when Facebook sent the cease and desist letter, Power, as it conceded, knew that it no longer had permission to access Facebook's computers at all. Power, therefore, knowingly accessed and without permission took, copied, and made use of Facebook's data."). The Supreme Court recently denied Power Ventures's petition for certiorari; *Power Ventures* would have provided the Court with its first opportunity to bridge the gulf between broad and narrow interpretations of 18 U.S.C. § 1030(a)(2)(C).

It is worth noting that CFAA claims can intersect with copyright claims, particularly those arising under the Digital Millennium Copyright Act. *See* Facebook, Inc. v. Power Ventures, Inc., No. C-08-5780 (N.D. Cal. Sept. 22, 2009) (order denying motion to dismiss and granting in part and denying in part motion for more definite statement) (plaintiff alleging violations of both the CFAA and DMCA). Currently, a company called hiQ is seeking declaratory judgment that its scraping of LinkedIn's website violates neither the CFAA nor the DMCA. *See* Complaint, hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. 2017) (No. 3:17-cv-03301), https://ia801501.us.archive.org/10/items/gov.uscourts.cand.312704/gov.uscourts.cand.312704.1.0.pdf [https://perma.cc/HX5E-HKY4]. In August 2017, Judge Chen granted hiQ's motion for a preliminary injunction that barred LinkedIn from blocking hiQ's data collection from the LinkedIn website. Order Granting Plaintiff's Motion for Preliminary Injunction, hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. 2017) (No. 3:17-cv-03301).

32. Under the CFAA, there is no exemption for accessing a computer system without authorization or in excess of operation for the purposes of news reporting, scholarship, or research. I believe that adopting a CFAA exemption that mirrors the fair use doctrine in copyright law would be valuable, particularly for examining and exposing bias in AI systems. *See infra* Part IV.

33. *See* Complaint, Sandvig v. Lynch, No. 1:16-cv-01368 (D.D.C. June 29, 2016). The CFAA would be well-served by judicial adoption of a flexible standard not unlike the copyright doctrine of fair use, discussed *infra* Part III.

34. Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. (forthcoming 2018) [hereinafter *Life, Liberty, and Trade Secrets*]; Rebecca Wexler, *When a Computer Program Keeps You in Jail*, N.Y. TIMES (June 13, 2017), https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html [https://perma.cc/BMW4-XPQ6]; *see also* Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. 102 (2017) (discussing how trade secret law

Copyright law causes friction that limits access to training data and restricts who can use certain data. This friction is a significant contributor to biased AI.[35] The friction caused by copyright law encourages AI creators to use biased, low-friction data (BLFD) for training AI systems, like the word2vec toolkit, despite those demonstrable biases.[36] As Google's decision not to freely release the Google News corpus reveals, copyright law can also curtail the implementation of bias mitigation techniques, including interventions like reweighting algorithmic inputs or supplementing datasets with additional data.[37] Copyright law can even preclude potential competitors from converting the customers of dominant AI players.

This Article is the first to examine the ways in which law channels AI in a fundamentally biased direction. In Part I, I briefly explain the mechanics of training AI to examine how bias is introduced into and embedded in AI systems. Part II explores the problem of bias through the lens of the copyright doctrine, looking at how the law's exclusion of access to certain copyrighted source materials may create or promote biased AI by constraining competition to create less biased AI. This Part also unpacks the biases embedded in two appealing sources of BLFD, public domain works and Creative Commons-licensed works. Part III examines how another area of copyright law, the fair use doctrine, has been used to address concerns of competition, access, and fairness in the context of other innovative computational technologies and ask whether

---

can be used to shield policing algorithms from public scrutiny); Sonia K. Katyal, *Algorithmic Civil Rights* (discussing the same and suggesting a whistleblowing framework to enable disclosure of biased algorithms) (draft manuscript on file with author).

35. *See* Amanda Levendowski, "Fair Use for Fairer AI," *Lightning Round: Private Law and Public Answers*, WE ROBOT (Apr. 1, 2017), https://yalelaw.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=818e6f21-3333-4fb7-91b7-780907aa74af [https://perma.cc/P6BD-Y64U] (discussing how copyright law channels AI in a fundamentally biased direction).

36. It is worth noting that ImageNet, one of the largest repositories of images used to train AI systems, disclaims copyright ownership in any of the images in its database and is structured to sidestep allegations of copyright infringement. *See About ImageNet*, IMAGENET (2016), http://image-net.org/about-overview [https://perma.cc/24MY-T4WX] ("No, ImageNet does not own the copyright of the images."); *Download FAQ*, IMAGENET (2016), http://image-net.org/download-faq [https://perma.cc/H236-U6P8] ("The images in their original resolutions may be subject to copyright, so we do not make them publicly available on our server."). ImageNet also positions itself as a search engine, which have long been held to be fair use. *Id.* ("ImageNet only provides thumbnails and URLs of images, in a way similar to what image search engines do. In other words, ImageNet compiles an accurate list of web images for each synset of WordNet."); Perfect 10, Inc. v. Amazon.com, 508 F.3d 1146 (9th Cir. 2007) (finding that search engines are fair use); Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003). Notably, ImageNet was used in the study in which an AI system adopted and amplified the implicit biases in images. *See* Zhao, *Men Also Like Shopping, supra* note 9 (examining gender bias contained in visual datasets).

37. *See Debiasing Word Embeddings*, *supra* note 4, at 3.

fair use is equally capable of addressing those concerns in the field of AI bias. Given that the normative values embedded in the tradition of fair use ultimately align with the goals of mitigating bias, I conclude that fair use is up to the task of, quite literally, promoting fairer AI.

I. TEACHING SYSTEMS TO BE ARTIFICIALLY INTELLIGENT

When journalists, researchers, and even engineers say "AI," they tend to be talking about machine learning, a field that blends mathematics, statistics, and computer science to create computer programs with the ability to improve through experience automatically.[38] This Part does not attempt to provide a comprehensive account of how AI is engineered or used in practice. Rather, it is intended to provide non-technical readers with an accessible introduction to the mechanics of training AI systems and to preview how bias can be introduced into and embedded in those systems.

Most AI systems are trained using vast amounts of data and, over time, hone the ability to suss out patterns that can help humans identify anomalies or make predictions.[39] Well-designed AI systems can automatically tweak their analyses of patterns in response to new data, which is why these systems are particularly useful for tasks that rely on principles that are difficult to explain, such as the organization of adverbs in English,[40] or when coding the program would be impossibly complicated.[41] AI systems are not especially new technology: the United States Postal Service began developing an AI system to decipher the rich

---

38. TOM M. MITCHELL, MACHINE LEARNING, 1–2 (1997). The term "machine learning" is attributed to Arthur L. Samuel, who coined the term in his seminal article about teaching a computer to play checkers. *See* Arthur L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. RES. & DEV. 535, 535 (1959), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.2254&rep=rep1&type=pdf [https://perma.cc/4S4Y-LCE2] [hereinafter *Some Studies in Machine Learning*]. This Article uses the broader term AI, though I am generally referring to machine learning techniques. To paraphrase James Grimmelmann, "that usage fight has already been lost." James Grimmelmann, *Copyright for Literate Robots*, 101 IOWA L. REV. 657, 658 n.1 (2016).

39. *See Some Studies in Machine Learning*, *supra* note 38, at 535–37.

40. In *The Elements of Eloquence*, for example, entomologist Mark Forsyth explains that "adjectives in English absolutely have to be in this order: opinion-size-age-shape-colour-origin-material-purpose Noun." MARK FORSYTH, THE ELEMENTS OF ELOQUENCE: SECRETS OF THE PERFECT TURN OF PHRASE 45 (2014). He notes that "every English speaker uses that list, but almost none of us could write it out." *Id.* at 46. Teaching such a construct would be an ideal use case for an AI system: a human would be hard-pressed to identify and articulate such a rule but an algorithm could parse out the rule for itself on review of many adverb-laden sentences. NLP thrives on these sorts of rules.

41. MITCHELL, *supra* note 38, at 1–5.

variations of human handwriting on letters back in 1992.[42] To develop an accurate AI system for decoding frenetically scribbled addresses, the system would need to be exposed to thousands of variations and examples on how humans write letters and numbers, then string the two together. Most AI systems require exposure to significant amounts of data to automatically improve at a task. These data are referred to as "training data."[43]

To illustrate the mechanics of teaching an AI system, take a low-stakes example: training an AI system to recognize a cat.[44] Researchers could manually go through each image in the training data and label a cat or specify a series of features that comprise a cat. Both approaches are examples of "supervised learning," the technique overwhelmingly used to train commercial AI systems.[45] By way of analogy, supervised

42. Tim Race, *Moving Scribbled Mail Along*, N.Y. TIMES (May 27, 1992), http://www.nytimes.com/1992/05/27/business/moving-scribbled-mail-along.html?pagewanted=all [https://perma.cc/WPD3-PKYA].

43. MITCHELL, *supra* note 38, at 4.

44. Identifying cats is frequently invoked as a use-case for AI systems. *See, e.g.*, Gideon Lewis-Kraus, *The Great A.I. Awakening*, N.Y. TIMES: MAG. (Dec. 14, 2016), https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html [https://perma.cc/R8UF-CLR2] ("Imagine you want to program a cat-recognizer on the old symbolic-A.I. model."). For alternate, excellent explanations of the mechanics of teaching AI systems from lawyers, see Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 88, 90–93 (2014) [hereinafter *Machine Learning and Law*]; David Lehr & Paul Ohm, *Playing with the Data*, https://www.law.upenn.edu/live/files/6551-ohm-lehr-playing-with-the-datapdf [https://perma.cc/PG7A-YZFJ] (forthcoming).

45. Alternately, researchers can set an AI system loose on training data with limited human guidance and leave it to the system to determine which features comprise the concept of a cat, a technique called "unsupervised learning." *See generally* H.B. Barlow, *Unsupervised Learning*, *in* UNSUPERVISED LEARNING: FOUNDATIONS OF NEURAL COMPUTATION (Geoffrey Hinton & Terrence J. Sejnowski eds., 1999). To continue the analogy, unsupervised learning would be like that same child wandering a neighborhood, gradually learning which creatures were cats for herself, without any parental guidance or intervention. Indeed, Google researchers notoriously exposed an AI system to unlabeled, still images derived from 10 million YouTube videos—and the system began identifying cats. QUOC V. LE ET AL., BUILDING HIGH-LEVEL FEATURES USING LARGE SCALE UNSUPERVISED LEARNING, NAT'L CONF. ON MACH. LEARNING (2012), https://static.googleusercontent.com/media/research.google.com/en//archive/unsupervised_icml2012.pdf [https://perma.cc/PW5U-MFCG]. The AI system produced an image that reflected its "perception" of a cat. Jeff Dean & Andrew Ng, *Using Large-Scale Brain Simulations for Machine Learning and A.I.*, GOOGLE BLOG: MACH. LEARNING (June 26, 2012), https://googleblog.blogspot.com/2012/06/using-large-scale-brain-simulatioreversens-for.html [https://perma.cc/5P7R-E8KB]. Unsupervised learning is often the topic of cutting-edge research and scholarship, but even Jeff Dean, one of the researchers involved with the Google Cat paper, has acknowledge that the limitations of unsupervised learning as it currently exists. Tom Simonite, *The Missing Link of Artificial Intelligence*, MIT TECH. REV. (Feb. 18, 2016), https://www.technologyreview.com/s/600819/the-missing-link-of-artificial-intelligence/ [https://perma.cc/76DG-RFPG].

It is worth noting that using YouTube videos for training AI systems would not be an option to any or all researchers or AI creators, as discussed *infra* Part II.

learning is not unlike a child walking around with a parent who constantly points out cats and affirms (or corrects) the child's own perceptions.

Good training data is crucial for creating accurate AI systems.[46] The AI system tasked with identifying cats must be able abstract out the right features, or heuristics, of a cat from training data.[47] To do so, the training data must be well-selected by humans—training data infused with implicit bias can result in skewed datasets that fuel both false positives and false negatives. For example, a dataset that features only cats with tortoiseshell markings runs the risk that the AI system will "learn" that a mélange of black, orange, and cream markings are a heuristic for identifying a cat and mistakenly identify other creatures, like brindle-colored dogs, as cats.[48] Similarly, a dataset that features only mainstream domestic cats could create an AI system that "learns" that cats have fluffy fur, pointy ears, and long tails and fail to identify cats of outlier breeds, like a Devon Rex, Scottish Fold, or Manx.[49] And, in both examples, all manner of wildcats are excluded from the training data.

It is understandable, perhaps even excusable, for an AI creator with anything less than an obsessive knowledge of cats to overlook how biased training data could "teach" an AI system heuristics that make it very difficult, if not impossible, to accurately perform the desired task. Teaching an AI system to recognize a cat can be decidedly low-stakes, but consider how errors rooted in biased training data play out when the stakes are raised. Take, for example, the Boston man who had his

---

46. Buolamwini & Gebru, *supra* note 23, at 1 ("It has recently been shown that algorithms trained with biased data have resulted in algorithmic discrimination.").

47. *Machine Learning and Law*, *supra* note 44, at 91 (discussing how an email spam-detection system would build heuristics for identifying and classifying spam).

48. In AI, a system that builds a heuristic too closely tied to biased training data demonstrates a problem called overgeneralization. *Machine Learning and Law*, *supra* note 44, at 106 ("The general idea is that it is undesirable for a machine learning algorithm to detect patterns in the training data that are so finely tuned to the idiosyncrasies or biases in the training set such that they are not predictive of future, novel scenarios.").

49. Devon Rexes have short curly fur, Scottish folds are known for their turned-down ears, and Manx cats have bobbed tails. *Devon Rex*, WIKIPEDIA (Frescobot: Apr. 7, 2018), https://en.wikipedia.org/w/index.php?title=Devon_Rex&oldid=835175125 [https://perma.cc/JVG3-GDFP]; *Scottish Fold*, WIKIPEDIA (RoySmith: Mar. 14, 2018), https://en.wikipedia.org/w/index.php?title=Scottish_Fold&oldid=830302374 [https://perma.cc/47QQ-RBE4]; *Manx cat*, WIKIPEDIA (Jabberjaw: Mar. 9, 2018), https://en.wikipedia.org/wiki/Manx_cat [https://perma.cc/TU7G-K797]. And many types of wildcats, including pumas, cougars, and bobcats, similarly depart from the aforementioned heuristics. *See, e.g.*, *Puma* (genus), WIKIPEDIA (Shellwood: Apr. 25, 2018), https://en.wikipedia.org/w/index.php?title=Puma_(genus)&oldid=838201736 [https://perma.cc/CV66-FCWL] (noting that the genus *Puma* includes cats among the largest felines in the cat family, weighing up to 220 pounds—much larger than the average domestic housecat).

driver's license revoked because an AI system mistakenly flagged him as a different fraudulent driver.[50] Or the Taiwanese engineering student who was briefly stranded in Australia when he was unable to renew his passport online because the AI system rejected his photo by incorrectly identified his eyes as being closed.[51] Or perhaps most troublingly, the commercial AI systems used by law enforcement agencies that are consistently less accurate for women, African Americans, and younger people.[52] Biased training data can play a role in all three errors.[53] When AI systems are increasingly used for purposes like these, the implicit biases resulting in Type 1 and Type 2 errors[54] become more than engineering goofs—they are dangerous.

II. COPYRIGHT LAW CAUSES FRICTION FOR CREATING FAIRER AI SYSTEMS

The internet may be full of cats, but it does not follow that the photographs and videos featuring those cats are free for anyone to use.[55] Behind the mechanics of training AI systems lurks the hairier matter of

---

50. Meghan E. Irons, *Caught in a Dragnet*, BOS. GLOBE (July 17, 2011), http://archive.boston.com/news/local/massachusetts/articles/2011/07/17/man_sues_registry_after_license_mistakenly_revoked/ [https://perma.cc/UZ7G-P8WY].

51. Cheng, *supra* note 24.

52. Scott J. Klum et al., *The FaceSketchID System: Matching Facial Composites to Mugshots*, 9 IEEE TRANSACTIONS ON INFO. FORENSICS & SEC. 12 (2014); Clare Garvie et al., *The Perpetual Line-Up: Unregulated Facial Recognition in America*, GEO. CTR. ON PRIVACY & TECH. (Oct. 18, 2016), https://www.perpetuallineup.org/ [https://perma.cc/8TBK-WS8K]. In 2016, the U.S. Government Accountability Office (GAO) issued a bluntly titled report, noting that the FBI does not even have a tolerable threshold for false positives, let alone a policy for testing these AI systems for bias. U.S. GOV'T ACCOUNTABILITY OFFICE, FACE RECOGNITION TECHNOLOGY: FBI SHOULD BETTER ENSURE PRIVACY AND ACCURACY (2016), https://www.gao.gov/assets/680/677285.pdf [https://perma.cc/3TD6-SHFP] (accessible version).

53. For a deeper discussion of training data and racial bias in facial recognition AI systems, see Alice O'Toole et al., *Face Recognition Algorithms and the "Other-Race" Effect*, 8 J. VISION 256 (2008), http://jov.arvojournals.org/article.aspx?articleid=2136933 [https://perma.cc/6E9C-UM2K].

54. A Type 1 error is also known as a false positive; a Type 2 error is also known as a false negative. *See Type I and Type II Errors*, WIKIPEDIA (Ed. Purgy Purgatorio, Jan. 1, 2018), https://en.wikipedia.org/w/index.php?title=Type_I_and_type_II_errors&oldid=818086578 [https://perma.cc/9G82-FEST].

55. *See Cats and the Internet*, WIKIPEDIA (Ed. Bender the Bot, Feb. 24, 2017), https://en.wikipedia.org/w/index.php?title=Cats_and_the_Internet&oldid=767149092 [https://perma.cc/CWJ7-66G3]; Jennifer A. Kingson, *'How Cats Took Over the Internet' at the Museum of the Moving Image*, N.Y. TIMES (Aug. 6, 2015), https://www.nytimes.com/2015/08/07/arts/design/how-cats-took-over-the-internet-at-the-museum-of-the-moving-image.html?_r=0 [https://perma.cc/D8JL-DGL5].

copyright protection.[56] Many AI systems are taught to "think" by reading, viewing, and listening to copies of human-created works.[57] These works, including books and articles, photographs, films and videos, and audio recordings, are not merely well-suited for use as training data for AI—they are also often protectable by copyright.[58]

The rules of copyright law grant exclusive rights to copyright owners, including the right to reproduce their works in copies, and one who violates one of those exclusive rights "is an infringer of copyright."[59] The current state-of-the-art in AI is in flux, but it seems inevitable that AI will be the latest computational technology to "pentest" the boundaries of copyright law.[60] Historically, copyright owners have regarded innovative computational technologies—from time-shifting home video recorders[61] to reverse-engineering software[62]—with skepticism (if not outright hostility) that invariably results in litigation.[63]

---

56. At least one famed internet feline, Tardar Sauce (better known as Grumpy Cat), has been involved in copyright litigation over the use of her image. *See* Complaint, Grumpy Cat Ltd. v. Grenade Beverage LLC, No. 8:15-cv-02063 (C.D. Cal. Dec. 11, 2015).

57. *See supra* Part I.

58. In the United States, copyright subsists in "original works of authorship fixed in any tangible medium of expression." 17 U.S.C. § 102(a) (2012). The same tends to be true for Berne Convention signatories. *See* Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886 (1971, amend. 1979) [hereinafter Berne Convention]. There are exceptions to this rule, discussed *infra* section II.B.1.

59. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 433 (1984) (paraphrasing 17 U.S.C. § 501(a)). This Article focuses on the reproduction right, which is but one example of an exclusive right that may be implicated when using copyrighted works as training data for AI systems. AI algorithms may implicate the right to prepare derivative works; training datasets may implicate the rights of display and performance. 17 U.S.C. § 106(3), (5)–(6). Given that the state-of-the-art in AI is rapidly evolving and that commercial players seek to keep proprietary algorithms and datasets private, as discussed in this Part and *infra* section III.D, it can be difficult to generalize about which other rights may be implicated and under what circumstances without resorting to speculation.

60. "Pentest" is shorthand for "penetration test," which refers to testing a system for weaknesses, limitations, and other vulnerabilities.

61. *Sony Corp. of Am.*, 464 U.S. 417 (1984) (copyright owners unsuccessfully suing manufacturer of home video recording technology for infringement).

62. *Compare* Sony Comp. Entm't, Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000) (copyright owner unsuccessfully suing manufacturers of reverse-engineered game cartridges for infringement), *and* Sega Enters. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992), *with* Atari Games Corp. v. Nintendo of Am., 975 F.2d 832 (Fed. Cir. 1992) (affirming preliminary injunction in favor of copyright owner seeking to enjoin Atari from exploiting reverse-engineered game program).

63. *See generally* Am. Broad. Co. v. Aereo, Inc., __ U.S. __, 134 S. Ct. 2493 (2014) (copyright owners successfully suing provider of internet-streamed broadcast television programming for infringement); Fox News Network v. TVEyes, Inc., 883 F.3d 169 (2d Cir. 2018) (copyright owner successfully suing operator of searchable audiovisual media archive for infringement); Capitol Records, LLC v. ReDigi Inc., 934 F. Supp. 2d 640 (S.D.N.Y. 2013), *appeal pending* 16-2321 (2d Cir. 2017) (copyright owners successfully suing operator of secondhand digital music marketplace

To date, no court has yet determined whether a copy made to train AI is a "copy" under the Copyright Act of 1976, let alone whether such a copy is infringement. It remains an open question whether copies created for purposes of training AI systems constitute "copies" under the Copyright Act, which defines "copies" as "material objects . . . in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."[64] Thus, certain "copies" may be so fleeting that they are not considered copies at all.[65] Google, for example, has developed a technique called federated learning, which localizes training data to the originating mobile device rather than copying data to a centralized server.[66] It remains far from settled that decentralized training data stored in random access memory (RAM) would not be considered "copies" under the Copyright Act.[67]

Courts have also yet to confront whether unauthorized copies made for training AI are necessarily infringing copies. Copying works, or parts of a work, that are not protected by copyright is not infringement.[68]

---

for infringement); Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015) (copyright owners unsuccessfully suing mass book-digitization projects for infringement); Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014); A.V. *ex rel*. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 640 (4th Cir. 2009) (copyright owners unsuccessfully suing operators of digital plagiarism-detection service for infringement); Perfect 10, Inc. v. Amazon.com, 508 F. 3d 1146 (9th Cir. 2007) (copyright owners unsuccessfully suing operators of internet search and retail websites for infringement); Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2002); A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1013 (9th Cir. 2001) (copyright owners successfully suing providers of peer-to-peer file sharing websites for infringement); Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 259 F. Supp. 2d 1029 (C.D. Cal. 2003), *aff'd*, 380 F.3d 1154 (9th Cir. 2004), *aff'd*, 545 U.S. 913 (2005). I was previously an associate at Kirkland & Ellis LLP, which represented Fox News Network in the *TVEyes* case. The NYU Technology Law and Policy Clinic filed amicus briefs on behalf of clients in both *ReDigi* and *TVEyes*, the latter of which was filed prior to my work with the clinic.

64. 17 U.S.C. § 101 (2012).

65. *See, e.g.*, Cartoon Network LP v. CSC Holdings, Inc., 536 F.3d 121, 130 (2d Cir. 2008) (holding that movies and television programs streamed through a data buffer for 1.2 seconds did not create copies under the Copyright Act). *But see* MAI Sys. Corp. v. Peak Computer, Inc., No. 92-1654, 1992 WL 159803 (C.D. Cal. Apr. 14, 1992), *aff'd*, 991 F.2d 511 (9th Cir. 1993) (holding that software stored in random access memory (RAM) constituted an infringing copy).

66. *See* H. BRENDAN MCMAHAN ET AL., arXiv:1602.05629, COMMUNICATION-EFFICIENT LEARNING OF DEEP NETWORKS FROM DECENTRALIZED DATA (2016), https://arxiv.org/abs/1602.05629 [https://perma.cc/GF9D-L8XE].

67. For thoughtful suggestion of a predictable, flexible approach to transitory "copies" that avoids the vulnerabilities of *Peak*, see Aaron Perzanowski, *Fixing Ram Copies*, 104 NW. U. L. REV. 1067 (2010); 17 U.S.C. § 117.

68. *See* 17 U.S.C. § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle,

Similarly, the fair use doctrine, discussed in Part III below, expressly states that certain uses of copyrighted works are "not an infringement of copyright."[69] Or copying may be so trivial, so de minimis, that copyright law does not concern itself with such copies.[70] Judge Pierre Leval has suggested that photocopying a *New Yorker* "cartoon to put on the refrigerator" belongs to a category of copyright "[q]uestions that never need to be answered."[71] If an answer were required, the de minimis doctrine would conclude that such copying is not infringement.[72] Copying many cartoons to train a neural network how to imitate the style of *New Yorker* cartoons could be similarly de minimis.[73] However, it still remains to be seen how that principle plays out when commercial AI systems abstract generic features from images to "learn" what makes a cat a cat.

If history is any indication, courts will soon be confronted with these and many other fact-specific questions about using copyrighted works as training data for AI systems.[74] When the cost of infringement can run as high as $150,000 for each infringing copy, few AI creators can afford to

---

or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.").

69. *Id.* § 107; *infra* Part III.

70. The doctrine gets its name from the aphorism *de minimis non curat lex*, meaning "[t]he law does not concern itself with trifles." BLACK'S LAW DICTIONARY 464 (8th ed. 2004). For that precise reason, case law reliant on the de minimis doctrine is rather sparse.

71. Pierre N. Leval, *Nimmer Lecture: Fair Use Rescued*, 44 UCLA L. REV. 1449, 1457 (1997). At least some courts have concurred with Judge Leval's hypothetical. *See, e.g.*, Ringgold v. Black Entm't Television, Inc., 126 F.3d 70, 77 (2d Cir. 1997) (holding that display of a poster depicting artist's quilt in the background of a television series for a total of "26 to 27 seconds" was "not *de minimis* copying").

72. *See, e.g.*, On Davis v. The Gap, Inc., 246 F.3d 152, 173 (2d Cir. 2001), as amended (May 15, 2001) ("Trivial copying is a significant part of modern life . . . . Because of the *de minimis* doctrine, in trivial instances of copying, we are in fact not breaking the law. If a copyright owner were to sue the makers of trivial copies, judgment would be for the defendants. The case would be dismissed because trivial copying is not an infringement.").

73. Such an AI system is not science fiction: researchers from the University of Tubingen have developed a neural algorithm capable of "painting" in the style of artistic masters like Vincent van Gogh and Edvard Munch. Notably, both masters' works are in the public domain and thus free of the friction created by copyright law; the underlying code for the project, however, is not publicly available. LEON A. GATYS ET AL., arXiv:1508.06576v2, A NEURAL ALGORITHM OF ARTISTIC STYLE (2015), https://arxiv.org/pdf/1508.06576v2.pdf [https://perma.cc/WL3A-HD5C].

74. *Supra* note 63 and accompanying text (discussing copyright infringement litigation involving innovative computational technologies). Indeed, some of these issues are raised in the complaint recently filed by hiQ Labs seeking declaratory judgment that scraping LinkedIn's website does not violate the CFAA or DMCA. Complaint, hiQ Labs, *supra* note 31. As discussed, data scraping is a common means of collecting works for use as AI training data. *Id.*

take a gamble.[75] Thus, the rules of copyright law can be understood as causing two kinds of friction: competition and access. From a competition perspective, copyright law can limit implementation of bias mitigation techniques on existing AI systems and constrain competition to create less biased systems. And from an access perspective, copyright law can privilege the use of certain works over others, inadvertently encouraging AI creators to use easily available, legally low-risk works as training data, even when those data are demonstrably biased.

This section examines the ways in which the friction caused by copyright law can create or promote biased AI systems. I begin examining how copyright law limits testing AI systems through reverse engineering and algorithmic accountability processes or competing to convert customers. I then turn to unpacking the biases embedded in two attractive sources of BLFD—public domain works and Creative Commons-licensed works—that are readily available to AI creators.

A. Limiting Meaningful Accountability and Competition

The fundamental purpose of copyright law is to "promote the Progress of Science and useful Arts."[76] In practice, the rules of copyright law massively favor incumbents by causing friction for others to implement bias mitigation techniques or compete to converting customers.[77] Because AI systems are trained to "think" by reading, viewing, and listening to copies of human-created works, many of which are protectable by copyright law, the ability to acquire legal access to those works can play a determinative role in which companies can effectively compete in the marketplace.

A skim of the seven dominant commercial AI creators—Apple, Baidu, DeepMind, Facebook, Google, IBM, and Microsoft—puts this

75. 17 U.S.C. § 504(c)(2) (2012) (explaining that a court may increase the award of statutory damages "to a sum not more than $150,000" for a finding of willful infringement, and that a court "may reduce the award of statutory damages to a sum of not less than $200" even if an infringer proves—and the court finds—that "the infringer was not aware and had no reason to believe his or her acts constituted an infringement of copyright").

76. U.S. CONST. art. I, § 8, cl. 8.

77. Incumbent companies are in both the best position to use copyrighted works as AI training data without detection and, if it came to it, the best position to defend themselves against allegations of infringement. *See* LAWRENCE LESSIG, FREE CULTURE 125 (2004); *id.* at 187 ("[F]air use in America simply means the right to hire a lawyer to defend your right to create. And as lawyers love to forget, our system for defending rights such as fair use is astonishingly bad—in practically every context, but especially here.").

dynamic in stark terms.[78] The AI playing field largely mirrors the major players in the technology sector and, perhaps unsurprisingly, reflects the same lack of diversity. For example, all but one of the companies are based in the United States.[79] Most have some of the greatest market capitalization rates in the world.[80] Two are owned by the same parent company.[81] Only one is led by a woman.[82] In specialized sectors, such as the AI systems used by law enforcement, even fewer companies dominate the market.[83] As journalist Tom Simonite recently put it, "when competition in tech depends on machine learning systems

78. Eric Jang, *What Companies Are Winning the Race for Artificial Intelligence?*, FORBES (Feb. 24, 2017), https://www.forbes.com/sites/quora/2017/02/24/what-companies-are-winning-the-race-for-artificial-intelligence/#28e2ebff5cd8 [https://perma.cc/7AGN-BCGU]. OpenAI, a nonprofit AI research company, is also often mentioned as a dominant player in AI. *Id.*; *see also About*, OPENAI, https://openai.com/about/ [https://perma.cc/ZS5D-JNBT].

79. Baidu is headquartered in Beijing, China. This lack of geographic diversity is, as least in some part, attributable to the European Union's strict privacy regulations—including the Database Directive General Data Protection Regulation (GDPR), which expressly limits data mining and profiling—that can inhibit development of AI systems. Directive 96/9/EC *and* GDPR, Art. 13(2)(f). For scholarship examining the potential effects of the GDPR on AI decisionmaking, see FINALE DOSHI-VELEZ & MASON KORTZ, arXiv:1711.01134v2, ACCOUNTABILITY OF AI UNDER THE LAW: THE ROLE OF EXPLANATION (2017), https://arxiv.org/pdf/1711.01134.pdf [https://perma.cc/Y649-4VTM]; Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise*, BERKELEY TECH. L.J. (forthcoming), https://poseidon01.ssrn.com/delivery.php?ID=488007067005081121094103080065080110058022049054058085108021085027104004025091114018122027044107111060037066096095074092108115006036025010060121126073021075071110071034005024009122100127114115115064004030065115023113105113071029003110075069116097078073&EXT=pdf [https://perma.cc/S7BV-NF3Z].

80. *See Fortune Global 500*, FORTUNE (2017), http://fortune.com/global500/list/filtered?sector=Technology [https://perma.cc/8MLZ-4USN] (identifying Alphabet, Apple, Facebook, IBM, and Microsoft).

81. Alphabet is currently the parent company of both DeepMind, which was acquired for more than $500 million in 2014, and Google. *The Company*, DEEPMIND, https://deepmind.com/about/ [https://perma.cc/KZ9N-A5N2] ("Having been acquired by Google in 2014, we are now part of the Alphabet group."); Amir Efrati, *Google Beat Facebook for DeepMind, Creates Ethics Board*, INFO. (Jan. 26, 2014, 7:26 PM), https://www.theinformation.com/articles/Google-beat-Facebook-For-DeepMind-Creates-Ethics-Board?shared=1a79c7e6517e8665 [https://perma.cc/3ALV-RDYF].

82. The current chairman and CEO of IBM is Ginni Rometty, who also happens to be the first woman to lead the company. Ginni Rometty, PUTTING SMART TO WORK, THINK 2018 (Mar. 20, 2018) ("I do not want to be known as the first woman CEO of IBM. I just want to be known as the CEO of IBM.").

83. Axon, the new AI-centered brand identity of longtime law enforcement contractor Taser, dominates in the field of police body cameras. *See* David Gelles, *Taser International Dominates the Police Bodycam Market*, N.Y. TIMES (July 12, 2016), https://www.nytimes.com/2016/07/13/business/taser-international-dominates-the-police-body-camera-market.html [https://perma.cc/E4F7-FF8C]; Elizabeth Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. ONLINE 101, 114–16 (2017) (discussing how Axon has effectively cornered the policy bodycam market).

powered by huge stockpiles of data, slaying a tech giant may be harder than ever."[84]

Understanding the internal workings of AI systems is crucial to pinpointing potential sources of bias and implementing bias mitigation techniques. Several dominant AI players, including Google, IBM, and Microsoft, have released some of their algorithms as open source.[85] Releasing underlying datasets is far less common.[86] It is plausible, if not probable, that dominant AI players create unauthorized copies of protectable works to use as training data for AI systems. On the one hand, these companies are in the best position to defend themselves against allegations of infringement. But on the other, copyright law effectively incentivizes companies to create "black box" systems which, as Professor Frank Pasquale has examined extensively, provide outputs without disclosing how those outputs were determined.[87] These systems obfuscate the mechanics of operation, including training data, in a metaphorical black box, in part because revealing the workings of the AI systems to the public could mean more than scrutiny—it could mean liability.[88]

Even so, researchers, journalists, and competitors have managed to pop the top off black box computational systems using algorithmic accountability processes. ProPublica's groundbreaking exposé on the black box algorithm behind Northpointe's COMPAS algorithm has quickly become a canonical example of using both techniques to reveal and interrogate bias.[89]

The COMPAS tool is one of the most popular algorithmic risk scores used to evaluate criminal defendants in the United States.[90] The

---

84. Tom Simonite, *AI and 'Enormous Data' Could Make Tech Giants Harder to Topple*, WIRED (July 13, 2017), https://www.wired.com/story/ai-and-enormous-data-could-make-tech-giants-harder-to-topple/ [https://perma.cc/89N5-P9L3].

85. *Id.*

86. Cade Metz, *Google Open-Sourcing Tensorflow Shows AI's Future Is Data*, WIRED (Nov. 16, 2015), https://www.wired.com/2015/11/google-open-sourcing-tensorflow-shows-ais-future-is-data-not-code/ [https://perma.cc/2VGG-8ZKP].

87. *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY: ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); Anupam Chander, *The Racist Algorithm*, 115 MICH. L. REV. 1023 (2017) (reviewing THE BLACK BOX SOCIETY).

88. *See* Citron & Pasquale, *supra* note 26; Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 119–20 (2014).

89. *See* Angwin et al., *supra* note 16.

90. Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [https://perma.cc/FNF5-LMG7] [hereinafter *How We Analyzed COMPAS*].

COMPAS tool uses criminal defendants' responses to a questionnaire, usually administered at booking, to generate a series of predictions about whether the person is likely to reoffend, delivered in the form of "risk scores."[91] Northpointe has touted that COMPAS and algorithmic tools like it can "minimize subjective personal biases that exist in pretrial decision-making" and "improve the placement of individuals for treatment and public safety, protect courts against legal scrutiny, and improve the allocation of resources."[92] Northpointe has described its algorithm as a "theory-guided instrument" rather than an AI system, but the code underlying the COMPAS tool has never been released.[93] Nevertheless, judges increasingly began to rely on COMPAS risk scores to inform life-altering decisions about bail, parole, and sentencing without exposing the COMPAS tool to public examination of whether the risk scores were biased.[94]

Enter ProPublica. In 2016, the nonprofit newsroom sought to assess the COMPAS tool "to discover the underlying accuracy of their recidivism algorithm and to test whether the algorithm was biased against certain groups."[95] ProPublica relied on public records requests to acquire COMPAS risk scores for more than 11,000 criminal defendants who were assessed before trial, all from Broward County, Florida.[96] ProPublica then reconstructed criminal histories and subsequent incarceration records for each individual using public databases.[97] Armed with COMPAS risk scores and a dataset built from those individuals' criminal records, ProPublica reverse engineered which characteristics caused the COMPAS algorithm to predict higher recidivism risk scores.[98]

---

91. *Id*.

92. THOMAS BLOMBERG ET AL., FLA. ST. UNIV., VALIDATION OF THE COMPAS RISK ASSESSMENT CLASSIFICATION INSTRUMENT (2010), http://criminology.fsu.edu/wp-content/uploads/Validation-of-the-COMPAS-Risk-Assessment-Classification-Instrument.pdf [https://perma.cc/ED5L-9KW5] (citing D. A. Andrews et al., *The Recent Past and Near Future of Risk and/or Need Assessment*, 52 CRIME & DELINQUENCY 7 (2006)).

93. Tim Brennan & William L. Oliver, *The Emergence of Machine Learning Techniques in Criminology*, 12 CRIMINOLOGY & PUB. POL'Y 551, 559 (2013), http://criminology.fsu.edu/wp-content/uploads/volume-12-issue-3.pdf [https://perma.cc/8D9H-NWJF].

94. Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1205–07 (2017); *see also Life, Liberty, and Trade Secrets*, *supra* note 34, at 12, 23.

95. *How We Analyzed COMPAS*, *supra* note 90.

96. *Id.*

97. *Id.*

98. *Id.*

ProPublica concluded that race and gender were predictive of higher scores—and those predictors were biased. Controlling for other factors, black defendants and female defendants were more likely to get higher risk scores than white or male defendants.[99] ProPublica also tested the accuracy of the COMPAS recidivism risk scores and found that the scores were only accurate about 64% of the time.[100] And, based on ProPublica's testing, the scores were also racist: the COMPAS algorithm was twice as likely to wrongly predict that black defendants would be arrested for additional crimes after being released as compared with white defendants. The opposite mistake was made with white defendants.[101] ProPublica journalists Julia Angwin, Jeff Larson, and their colleagues went on to publish an investigative report detailing their findings and released both their statistical methodology and the dataset ProPublica assembled for use in its testing.[102]

Bias mitigation techniques, like reverse engineering and algorithmic accountability processes, provide a means of identifying where competitors may be able to make gains over incumbents: by rectifying a known bias. However, ProPublica's investigation shed light on biases embedded in COMPAS risk assessments and Northpointe continues to maintain active contracts across the country.[103]

---

99. *Id.*

100. *Id.*

101. *Id.*

102. *See* Angwin et al., *supra* note 16; *How We Analyzed COMPAS*, *supra* note 90. ProPublica also made the code underlying their analysis publicly available. *Data and Analysis for 'Machine Bias,'* GITHUB (June 12, 2017), https://github.com/propublica/compas-analysis [https://perma.cc/6A6D-Z5WR]. ProPublica's reporting was not without controversy. Researchers and Northpointe itself have disputed ProPublica's methodology and findings. *See* WILLIAM DIETERICH ET AL., NORTHPOINTE, COMPAS RISK SCALES: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY (2016), https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html [https://perma.cc/DG4V-MLEU]. ProPublica has publicly responded to both sets of critiques from Northpointe. Julia Angwin & Jeff Larson, *ProPublica Responds to Company's Critique of Machine Bias Story*, PROPUBLICA (July 29, 2016), https://www.propublica.org/article/propublica-responds-to-companys-critique-of-machine-bias-story [http://perma.cc/UX2K-GS43]; Jeff Larson & Julia Angwin, *Technical Response to Northpointe*, PROPUBLICA (July 29, 2016), https://www.propublica.org/article/technical-response-to-northpointe [https://perma.cc/K34E-L5D9].

103. MICH. DEP'T OF CORR., ADMINISTRATION AND USE OF COMPAS IN THE PRESENTENCE INVESTIGATION REPORT (2017), https://www.michbar.org/file/news/releases/archives17/COMPAS-at-PSI-Manual-2-27-17-Combined.pdf [https://perma.cc/S9UG-UWSZ]; NORTHPOINTE, COMPAS RISK & NEED ASSESSMENT SYSTEM: SELECTED QUESTIONS POSED BY INQUIRING AGENCIES (2012), http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf [https://perma.cc/BF9U-KL7U]; VA. DEP'T OF CORR., REENTRY PLANNING (2017), https://vadoc.virginia.gov/about/procedures/documents/800/820-2.pdf [https://perma.cc/MJ8U-SCMD];

Copyright law creates a system that overwhelmingly favors dominant players in AI in three key ways: chilling reverse engineering, restricting algorithmic accountability processes, and hindering meaningful competition to convert customers. I discuss each in turn, and examine why the friction caused by copyright law all but guarantees that our AI systems can be no less biased than the handful of companies that create them.[104]

1. Chilling Reverse Engineering

Reverse engineering is a way of leveraging available inputs or outputs to understand the mechanics of what happens inside a black box system.[105] The concept of reverse engineering is not new for copyright law.

In 1988, Congress extended copyright protection to code.[106] Computer software creators and computers soon realized that interoperability between computer systems and other software would be key to successful technology.[107] If a manufacturer was not willing to license the rights to develop compatible programs, however, creators turned to reverse engineering systems to understand what was going on in the black box.[108]

The burgeoning video game industry of the early 1990s kicked the tension between newcomers' desire for interoperability and incumbents' desire to retain their dominance into overdrive. Both Nintendo and Sony, which collectively held 50% of the market share for video games at the time,[109] sued companies seeking to develop interoperable game cartridges for copyright infringement over the interim copies of code necessary to understand the functional, nonprotectable elements of the

---

COMPAS, WIS. DEP'T OF CORR., https://doc.wi.gov/Pages/AboutDOC/COMPAS.aspx [https://perma.cc/8BUT-67YG].

104. *See supra* notes 19–22.

105. *See* Andy Greenberg, *How to Steal an AI*, WIRED (Sept. 30, 2016), https://www.wired.com/2016/09/how-to-steal-an-ai/ [https://perma.cc/Y6XK-YDCW] (citing FLORIAN TRAMÈR ET AL., arXiv:1609.02943, STEALING MACHINE LEARNING MODELS VIA PREDICTION APIS, PROCEEDINGS OF USENIX SECURITY (2016), https://arxiv.org/abs/1609.02943 [https://perma.cc/EX26-CMDU] (describing adversarial analyses that aim to assist in reverse engineering machine learning models)).

106. 17 U.S.C. §§ 102, 117 (2012).

107. Julie E. Cohen, *Reverse Engineering and the Rise of Electronic Vigilantism: The Intellectual Property Implications of "Lock-Out" Programs*, 68 S. CAL. L. REV. 1091, 1093 (1995).

108. *Id.*

109. *Id.* (citing Merrill Goozner, *Rivals Nose in on Nintendo*, CHI. TRIB., June 12, 1994).

program.[110] In *Sega Enterprises v. Accolade*,[111] the Ninth Circuit determined that, while the newcomer created copies of Sega's copyrightable code while attempting to reverse engineer the system, those copies were a fair use.[112] As Judge Reinhardt observed, if such copying was "*per se* an unfair use, the owner of the copyright gains a *de facto* monopoly over the functional aspects of his work—aspects that were expressly denied copyright protection by Congress."[113]

Reverse engineering is not, however, a free-and-clear way of testing AI systems for bias. Just over a decade after *Sega*, Congress enacted the Digital Millennium Copyright Act (DMCA).[114] The DMCA is best known for creating a notice-and-takedown framework for user-generated content,[115] but section 1201 also creates liability for circumventing—through hacking or some other means—a technological measure that "effectively controls access to a work" protected under copyright law.[116] Every three years, the Librarian of Congress is empowered to grant temporary exemptions from the anti-circumvention provisions of section

---

110. *Compare* Sega Enters., Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992) (holding that reverse engineering was fair use), *with* Atari Games Corp. v. Nintendo of Am., 975 F.2d 832 (Fed. Cir. 1992) (holding that reverse engineering was not an entirely fair use).

111. 977 F.2d 1510 (9th Cir. 1992).

112. For a discussion of the fair use doctrine, see *infra* Part III.

113. *Sega Enters*., 977 F.2d at 1526. Trade secret laws also permit the reverse engineering of code that might otherwise be protectable as a trade secret. *See Life, Liberty, and Trade Secrets*, *supra* note 34; Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181 (2017).

114. Pub. L. No. 105-304, 112 Stat. 2860 (1998) (codified as amended, in scattered sections of 17 and 28 U.S.C.); *see also* Perel & Elkin-Koren, *supra* note 113.

115. 17 U.S.C. § 512 (2012).

116. *Id.* § 1201. Digital rights management (DRM) technologies have become a widely-adopted way to take advantage of the DMCA anti-circumvention provisions. For a discussion of the politics of copyright protection technologies, and referring to those technologies as "digital rights management," see Pamela Samuelson, *DRM {and, or, vs.} the Law*, 46 COMM. ACM 41 (2003); Pamela Samuelson & Jason Schultz, *Regulating Digital Rights Management Technologies: Should Copyright Owners Have to Give Notice About DRM Restrictions?*, 6 J. ON TELECOM. & HIGH TECH. L. 41 (2007), http://people.ischool.berkeley.edu/~pam/papers/notice%20of%20DRM-701.pdf [https://perma.cc/8PGH-8ZPW].

Section 1201 has also been invoked, albeit unsuccessfully, to prevent competitors from creating interoperable or compatible products. *See, e.g.*, Lexmark Int'l, Inc. v. Static Control Components, Inc., 387 F.3d 522, 549 (6th Cir. 2005) (creator of microchipped refillable printer cartridges unsuccessfully suing third-party manufacturer of compatible cartridges for violating section 1201) ("Nowhere in its deliberations over the DMCA did Congress express an interest in creating liability for the circumvention of technological measures designed to prevent consumers from using consumer goods while leaving the copyrightable content of a work unprotected."); Chamberlain Grp., Inc. v. Skylink Techs., Inc., 381 F.3d 1178 (Fed. Cir. 2004) (creator of garage door openers equipped with rolling codes unsuccessfully suing third-party manufacturer of garage door openers).

1201.[117] Currently, some forms of reverse engineering are exempt,[118] as well as encryption research and security testing.[119]

These exemptions have enabled researchers to tinker with the code in cars and expose vulnerabilities in voting machines.[120] To date, there has never been a petition to the Librarian of Congress requesting an exemption for testing AI systems for bias.[121]

Reverse engineering can be a critical means of examining bias in AI systems. However, the cost of getting legal advice to lawfully reverse engineer a system, coupled with the costs of getting it wrong, can chill researchers, journalists, and competitors from examining biases in AI systems effectively. Indeed, these steep costs have led journalism professor Nicholas Diakonpolous to conclude that "non-professional journalists may find it more difficult to do algorithmic-accountability investigations."[122]

---

117. 17 U.S.C. § 1201(a)(1).

118. *Id.* § 1201(f) (exempting reverse engineering "for the sole purpose of identifying and analyzing those elements of the program that are necessary to achieve interoperability of an independently created computer program with other programs"). Activities by law enforcement, intelligence agents, and other government activities are notably exempt. *See id.* § 1201(e).

119. *Id.* § 1201(g) (exempting encryption research "necessary to identify and analyze flaws and vulnerabilities of encryption technologies applied to copyrighted works, if these activities are conducted to advance the state of knowledge in the field of encryption technology or to assist in the development of encryption products"); *id.* § 1201 (exempting authorized security testing of a "computer, computer system, or computer network, solely for the purpose of good faith testing, investigating, or correcting, a security flaw or vulnerability").

120. Legal scholar Andrea Matwyshyn was integral in advocating for the exemptions that enabled testing vulnerabilities in voting machines. *See* Letter from Andrea Matwyshyn to the Honorable Jacqueline C. Charlesworth, General Counsel and Associate Register of Copyrights, Re: Docket No. 2014-7 Exemptions to Prohibition Against Circumvention of Technological Measures Protecting Copyrighted Works (June 29, 2015), https://www.copyright.gov/1201/2015/post-hearing/answers/Class_25_Hearing_Response_Matwyshyn_et_al_Docket_No_2014-07_2015.pdf [https://perma.cc/Q86A-2RW2]; Barb Darrow, *How Hackers Broke into U.S. Voting Machines in Less Than 2 Hours*, FORTUNE (July 31, 2017), http://fortune.com/2017/07/31/defcon-hackers-us-voting-machines/ [https://perma.cc/R6PZ-UGM3] (recounting that DEF CON attendees managed to compromise thirty different voting machines in less than two hours).

121. The Copyright Office is currently in the midst of its section 1201 rulemaking process, but no such exemption is pending. *See Seventh Triennial Section 1201 Proceeding (2018)*, U.S. COPYRIGHT OFFICE, https://www.copyright.gov/1201/2018/ [https://perma.cc/W5T9-J5BF]. It is my hope to work with academics, researchers, journalists, and other stakeholders to craft a proposed section 1201 exemption for testing AI systems during the next rulemaking in 2021.

122. Nicholas Diakonpolous, *Algorithmic Accountability*, 3 DIGITAL JOURNALISM 398, 410 (2015), http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/algorithmic_accountability_final.pdf [https://perma.cc/B2RE-NLCU] (specifically addressing the reverse engineering penalties posed by the Digital Millennium Copyright Act and Computer Fraud and Abuse Act); *see also* Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016).

2. Restricting Algorithmic Accountability Processes

Algorithmic accountability aims to bring values like transparency, explainability, and oversight to the development and deployment of AI systems. Journalistic reporting, like ProPublica's investigation into Northpointe's COMPAS tool, is one means of algorithmic accountability. Relying on whistleblowers to disclose the biases of AI systems may be another way.[123] Crowdsourcing audits of AI systems is yet another.[124] While these may seem like disparate approaches, the rules of copyright law can constrain all three.

The ProPublica investigation relied heavily on responses to public records requests to obtain the outputs, namely the COMPAS risk scores, and inputs, like criminal histories and subsequent incarceration records, necessary to reverse engineer aspects of the algorithm.[125] As part of its algorithmic accountability reporting, ProPublica also published the responses to its record requests publicly. Setting aside the logistics of public records requests, imagine instead that ProPublica had reverse engineered aspects of an algorithm to determine that the dataset was scraped from a proprietary, but publicly viewable, database of photographs. Even if ProPublica discovered that the works used as training data had biased the AI system, copying and publishing the photographs as a dataset could attract legal threats of infringement.

Whistleblowing poses a similar challenge, as an internal employee seeking to expose biased AI training data or algorithms would almost certainly have to copy the data before sharing it. Professor Sonia Katyal has suggested adapting the whistleblower-immunity regime to protect those who expose racist, sexist, and otherwise biased algorithms, but a federal system that formalizes protection for algorithmic whistleblowers does not yet exist.[126]

Examining biases in AI systems through techniques like reverse engineering is one step toward taking these systems to task; exposing those biases through algorithmic accountability investigations and reporting is a critical next step.[127] Copyright law, however, restricts the

---

123. Katyal, *supra* note 34.

124. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 211 (2016).

125. *How We Analyzed COMPAS*, *supra* note 90; *Data and Analysis for 'Machine Bias,' supra* note 102.

126. Katyal, *supra* note 34. As discussed previously, the Computer Fraud and Abuse Act is another law from which liability for whistleblowing could stem. *Supra* note 31.

127. The problem of training data and code non-disclosure are also prevalent in academia: one survey found that fewer 30% of the 400 papers presented at major conferences disclosed their

kinds of exposure in which journalists, researchers, and even competitors may feel comfortable engaging.

3. Hindering Competition to Convert Customers

It is no coincidence that dominant AI creators share not only status as global technology companies and the ensuing lack of diversity, but something else as well: they are masters of large-scale data acquisition. Acquiring copyrighted works to use as training data for AI systems can be exhaustingly resource intensive, but there are two ways to acquire those works without worrying about the threat of copyright infringement: AI creators can build a system to get those works[128] or buy them from someone else.

Facebook has mastered the "build-it" model by amassing training data from users in exchange for a service those users want, an approach that Professor Kathy Strandburg describes as "acquisition as a byproduct of another activity."[129] Facebook offers its social networking service to nearly 2 billion users, who are constantly creating and uploading massive numbers of messages and selfies that Facebook uses to train its AI systems.[130]

Facebook deploys its users' data to calibrate newsfeeds, generate alternate text for visually-impaired users, and create facial-recognition

---

training datasets, and a mere 6% include the underlying algorithmic code. *See* Matthew Hutson, *Missing Data Hinder Replication of Artificial Intelligence Studies*, SCIENCE (Feb. 15, 2018), https://www.sciencemag.org/news/2018/02/missing-data-hinder-replication-artificial-intelligence-studies [https://perma.cc/MF54-9X3C].

128. During a talk at the Stanford, Baidu's chief scientist Andrew Ng restated the build-it model: "[a]t large [tech] companies, we often launch products not for the revenue, but for the data . . . and we monetize the data through a different product." Evgeny Morozov (@evgenymorozov), TWITTER (Jan. 20, 2018, 12:40 AM), https://twitter.com/evgenymorozov/status/954634817198546945?lang=en [https://perma.cc/DFW3-4RJD] (citing Stanford Graduate Sch. of Bus., *Andrew Ng: Artificial Intelligence Is the New Electricity* at 33:27, YOUTUBE (Jan. 25, 2017), https://youtu.be/21EiKfQYZXc (last visited Apr. 26, 2018)).

129. Katherine J. Strandburg, *Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 5 (Julia Lane et al. eds., 2014), http://wpressutexas.net/cs378h/images/b/b3/LaneEtAlPrivacyBigDataAndThePublicGood.pdf [https://perma.cc/L9M9-RC4T].

130. *Company Info*, FACEBOOK, http://newsroom.fb.com/company-info/ [https://perma.cc/57PM-EVG5] (2.13 billion monthly active users as of Dec. 31, 2017); Dave Gershgorn, *Inside Facebook's Artificial Intelligence Lab*, POPULAR SCI. (Sept. 22, 2015), https://www.popsci.com/facebook-ai [https://perma.cc/3JSZ-DXAH]. Some of these creative works are almost certainly protectable by copyright law, but Facebook's Terms of Service and Privacy Policy essentially license the information from users to "[p]rovide, [i]mprove, and develop [s]ervices." *Data Policy*, FACEBOOK, https://www.facebook.com/about/privacy [https://perma.cc/BBG9-N2E2].

algorithms that are nearly as accurate as human perception.[131] But Facebook's growing presence in the AI space does not come cheap. It is a multi-billion-dollar corporation that employs nearly 200 individuals in its Fair AI Research arm alone, along with hundreds of other engineers and designers in machine learning, computer vision, and natural language processing (NLP).[132] NLP techniques can help AI creators streamline and improve language-centered AI systems, from translation and text prediction to search results and conversational chatbots. But to do so effectively, NLP algorithms rely on tremendous amounts of human-generated data to "learn" how humans use the written word. Inevitably, Facebook and other AI systems reliant on build-it models will reflect the biases of those systems' userbase. For example, although nearly 80% of American internet users also use Facebook,[133] the service has markedly less market penetration in the Middle East and Africa, making it difficult to create AI systems that adequately represent those users' experiences.[134]

IBM, on the other hand, has excelled at the "buy-it" model by getting works to use as AI training data through partnerships and acquisitions.[135]

---

131. *See, e.g.*, Julie Schiller & Omid Farivar, *Accessibility Research: Developing Automatic-Alt Text for Facebook Screen Reader Users*, FACEBOOK (Feb. 27, 2017), https://research.fb.com/accessibility-research-developing-automatic-alt-text-for-facebook-screen-reader-users/ [https://perma.cc/2RSK-NV9J] (discussing Facebook's Automatic Alt-Text service); Yaniv Taigman et al., *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*, *in* 2014 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 1705 (2014), http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6909616 [https://perma.cc/HS5D-FV4W] (explaining how Facebook's "DeepFace" AI closed in on a facial verification accuracy of 97.35% identification by learning from "a large collection of photos from *Facebook*").

132. *See People*, FACEBOOK, https://research.fb.com/people/page/4/?letter&cat=13 [https://perma.cc/6D6C-Y2PT].

133. Shannon Greenwood et al., *Social Media Update 2016*, PEW RES. CTR. (Nov. 11, 2016), http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [https://perma.cc/KM9C-EJYC].

134. *Id.* Facebook's attempts to penetrate these and other underserved markets to collect data on additional people in exchange for providing free, limited internet has been critiqued as "data colonialism." *See* Olivia Solon, *'It's Digital Colonialism': How Facebook's Free Internet Service Has Failed Its Users*, GUARDIAN (July 27, 2017), https://www.theguardian.com/technology/2017/jul/27/facebook-free-basics-developing-markets [https://perma.cc/Y5FX-6FS5]; Adrienne LaFrance, *Facebook and the New Colonialism*, ATLANTIC (Feb. 11, 2016), https://www.theatlantic.com/technology/archive/2016/02/facebook-and-the-new-colonialism/462393/ [https://perma.cc/MZS3-N7V]. For an in-depth accounting of the concept of data colonialism, see Jim Thatcher & David O'Sullivan, *Data Colonialism Through Accumulation by Dispossession: New Metaphors for Daily Data*, 34 ENV'T & PLAN. D: SOC'Y & SPACE 990 (2016), http://journals.sagepub.com/doi/abs/10.1177/0263775816633195 (last visited Apr. 26, 2018).

135. Cade Metz, *Google Open-Sourcing Tensorflow Shows AI's Future Is Data*, WIRED (Nov. 16, 2015), https://www.wired.com/2015/11/google-open-sourcing-tensorflow-shows-ais-future-is-data-

For example, IBM's Watson for Oncology teamed with Memorial Sloan-Kettering Cancer Center (MSK) to gain access to more than 12 million pages of medical literature, MSK "curated literature and rationale," and patient case histories, much of which could be protectable by copyright law.[136] The financial terms of the partnership were not disclosed publicly, but another IBM oncology partnership was a $50 million undertaking.[137] Similarly, IBM has purchased many smaller companies, along with those companies' valuable data. Recently, for example, IBM acquired a company called AlchemyAPI, which specialized in NLP and computer vision techniques.[138] AlchemyAPI had made a name for itself rather quickly. The company's NLP API processed 500,000 transactions within its first month.[139] It was heralded as making NLP techniques widely available,[140] and it successfully raised $2 million in seed funding.[141] The terms of IBM's deal were also not publicly disclosed,

---

not-code/ [https://perma.cc/2VGG-8ZKP] (discussing how IBM paid millions of dollars to acquire The Weather Channel "largely to acquire data it could use to feed its AI ambitions").

136. *See IBM Watson for Oncology*, IBM, https://www.ibm.com/watson/health/oncology/ [https://perma.cc/XRY3-VWY8]; *Memorial Sloan Kettering Trains IBM Watson to Help Doctors Make Better Cancer Treatment Choices*, MEM'L SLOAN KETTERING CANCER CTR. (Apr. 11, 2014), https://www.mskcc.org/blog/msk-trains-ibm-watson-help-doctors-make-better-treatment-choices [https://perma.cc/EF83-SSKD] (IBM and MSK note that "[a]ll identifying patient information is removed" prior to sharing the data with to Watson for Oncology) MSK clinicians and analysts also spent more than 15,000 hours training Watson on how to extract and interpret clinical research and physician notes. Laura Lorenzetti, *Here's How IBM Watson Health Is Transforming the Health Care Industry*, FORTUNE (Apr. 15, 2016), http://fortune.com/ibm-watson-health-business-strategy/ [https://perma.cc/9UGG-2D9L]. For now, I will set aside the potential for bias introduced by letting a partner that funds and publishes medical research train an AI system used to inform medical diagnoses and treatments.

137. IBM partnered with the Broad Institute of MIT and Harvard for access to genome data from thousands of drug-resistant tumors and cell-line studies. *IBM Watson Health and Broad Institute Launch Major Research Initiative to Study Why Cancers Become Drug Resistant*, IBM (Nov. 10, 2016), http://www.ibm.com/press/us/en/pressrelease/51032.wss [https://perma.cc/MC6Q-W58V]. Notably, IBM's high-friction data will become proprietary open data, discussed below, as the company has committed to making the data available publicly for research.

138. Press Release, IBM, IBM Acquires AlchemyAPI, Enhancing Watson's Deep Learning Capabilities (Mar. 4, 2015), https://www-03.ibm.com/press/us/en/pressrelease/46205.wss [https://perma.cc/YT2T-Z33R].

139. *Natural Language Understanding Demo*, IBM, https://alchemy-language-demo.mybluemix.net/ [http://perma.cc/YWH5-RAFB].

140. John de Goes, *"Big Data" Is Dead. What's Next?*, VENTURE BEAT (Feb. 22, 2013), https://venturebeat.com/2013/02/22/big-data-is-dead-whats-next/ [https://perma.cc/P9R2-TM2N].

141. Ron Miller, *IBM Watson Group Buys AlchemyAI to Enhance Machine Learning Capabilities*, TECHCRUNCH (Mar. 4, 2015), https://techcrunch.com/2015/03/04/ibm-watson-group-buys-alchemyapi-to-give-it-machine-learning-capabilities/ [https://perma.cc/NY9G-AQFU].

but AlchemyAPI's AI was recently rolled into IBM Watson products, including Watson Visual Recognition.[142]

But that is the rub. Imagine that a journalist manages to reverse engineer facial-detection algorithms created by dominant AI players using the build-it or buy-it models, discovers that the AI systems consistently struggle to detect the faces of Asian women, and publishes an in-depth report on these systems and their biases as a means of algorithmic accountability. A newcomer may be motivated to create an AI system without the race and gender biases of systems from the incumbent AI creators. But a newcomer would find it nearly impossible to build something approaching the vastness of Facebook's build-it model. It is equally unlikely that said newcomer could strike a licensing deal comparable to Google's agreement with global news agencies or a partnership equivalent to IBM's buy-it model.[143] Without the resources to get the vast amounts data easily acquired by major AI players, meaningful competition becomes all but nonexistent. Indeed, even a small company that manages to excel in the AI space is unlikely to be

142. *IBM Watson—AlchemyAPI*, IBM, https://www.ibm.com/watson/alchemy-api.html [https://perma.cc/P2HL-MMDC]; Melissa Mahoney, *A Unified Vision API*, IBM: CLOUD BLOG (May 19, 2016), https://www.ibm.com/blogs/bluemix/2016/05/alchemy-and-watson-visual-recognition-api/ [https://perma.cc/2U5D-ZLX9]; IBM has been notably mum about the exact sources of training data for Watson Visual Recognition, with one engineer stating that the algorithm is trained using "an IBM-curated (and ever-expanding) collection of image data from a wide variety of sources." Matthew Hill, *developerWorks Answers*, IBM (Mar. 14, 2016) https://developer.ibm.com/answers/questions/258248/what-are-the-image-databases-used-to-train-watson.html [https://perma.cc/N6Y7-72VL]. Watson also "memorized" the entirety of the crowd-sourced slang repository Urban Dictionary before it was purged from Watson's dataset for being a bit *too* colloquial. Alexis C. Madrigal, *IBM's Watson Memorizes the Entire "Urban Dictionary," then His Overlords Had to Delete It*, ATLANTIC (Jan. 10, 2013), https://www.theatlantic.com/technology/archive/2013/01/ibms-watson-memorized-the-entire-urban-dictionary-then-his-overlords-had-to-delete-it/267047/ [https://perma.cc/Q5QB-QABR].

143. Cade Metz, *Google Open-Sourcing Tensorflow Shows AI's Future Is Data*, WIRED (Nov. 16, 2015), https://www.wired.com/2015/11/google-open-sourcing-tensorflow-shows-ais-future-is-data-not-code/ [https://perma.cc/2VGG-8ZKP]. Indeed, Judge Denny Chin rejected the initially proposed Google Books settlement, which would have released Google from allegations of future acts of copyright infringement, because it was not "fair, adequate, and reasonable." Authors Guild v. Google, 770 F. Supp. 2d 666, 682 (S.D.N.Y. 2011). Part of Judge Chin's reasoning, however, was attributable to anti-competitive concerns raised by effectively granting Google a "de facto monopoly over unclaimed [orphan] works" and giving Google "control over the search market." *Id.* (referencing amicus briefs filed by the United States Department of Justice, as well as Amazon and Microsoft, raising antitrust concerns). Judge Chin also credited amici's privacy concerns regarding Google's potentially unlimited collection of information about readers' interests and habits, though it was not the focus of his opinion. *Id.* at 683–84. Notably, the Second Circuit came to a different conclusion when approving the *Literary Works* settlement. James Grimmelmann, *A Brief Note on Future Infringement*, LABORATORIUM (Aug. 18, 2011), http://laboratorium.net/archive/2011/08/18/a_brief_note_on_future_infringement [https://perma.cc/ZQ2T-FHX3].

competition for long—it remains more likely that the company will be snapped up by one of the dominant AI creators rather than competing against them.

Copyright law favors incumbent AI creators whose systems rely on high-friction build-it or buy-it models for acquiring training data. In so doing, the rules of copyright law cause friction for newcomers competing to convert customers. However, there is another approach to acquiring AI training data: the newcomer could use BLFD.

B. Privileging the Use of Biased, Low-Friction Data

Given the friction copyright law causes for accessing certain works, many AI creators turn to easily available, legally low-risk works to serve as training data for AI systems. Data derived from these works are often demonstrably biased—I call these biased, low-friction data (BLFD).[144] This is not to say that acquiring BLFD is easy or effortless, but rather to reflect that copyright law allows these works to be made so accessible that their use as training data is all but inevitable. The quintessential example of BLFD is a familiar to computer scientists: the Enron emails.

The "Enron emails" refer to the 1.6 million emails sent among Enron employees uploaded by Federal Energy Regulatory Commission (FERC) in 2003.[145] These emails remain one of the only large set of emails exchanged between real people in the world. The Enron emails are often colloquially referred to as being in the "public domain," but that is a legal misstatement.[146] While the Enron emails are available online publicly, they are more like orphan works: using the works still carries some risk, as getting permission from each of the authors is highly unlikely, but the comparative likelihood of a copyright infringement lawsuit is perhaps even more unlikely. The effect is that the Enron emails are perceived as posing an infinitesimally low legal risk because, though some of the Enron emails are protectable under copyright law, the practical likelihood of former Enron employees suing for copyright

144. The effects of friction on access and use of copyrighted works can be seen in another context: news consumption. Pay-walled news sources, such as the *The New York Times* and *The Washington Post*, will struggle to achieve the same impact as a website with free content, such as *Breitbart*.

145. Jessica Leber, *The Immortal Life of the Enron E-mails*, MIT TECH. REV. (July 2, 2013), https://www.technologyreview.com/s/515801/the-immortal-life-of-the-enron-e-mails/ [https://perma.cc/T9KM-SXCG]; Magalie R. Salas, *Third Order on Re-Release of Data Removed from Public Accessibility on April 7, 2003*, FED. ENERGY REG. COMM'N (Mar. 8, 2004), https://www.ferc.gov/whats-new/comm-meet/030304/E-46.pdf.

146. Leber, *supra* note 145.

infringement is exceedingly remote.[147] As such, the Enron emails and training corpora derived from them are freely available online in machine-readable formats.[148] The Enron emails are, as far as AI creators are concerned, as low-friction as it comes.

The Enron emails are ideal for teaching some types of AI, like email spam filters and folder systems,[149] but it is worth reminding ourselves why the Enron emails were released in the first place.[150] If you think there might be significant biases embedded in emails sent among employees of Texas oil-and-gas company that collapsed under federal investigation for fraud stemming from systemic, institutionalized unethical culture, you would be right. The Enron emails are simply not representative—not geographically, not socioeconomically, not even in terms of race or gender. Indeed, researchers have used the Enron emails *specifically* to analyze gender bias and power dynamics.[151] And yet the Enron emails remain a go-to dataset for training AI systems.[152]

The rules of copyright law that privilege BLFD as AI training data also have implications for privacy. The wealth of personal information used in both the build-it and buy-it models has created what Professor Julie Cohen calls "the biopolitical public domain: a repository of materials that are there for the taking and that are framed as inputs to particular types of productive activity."[153] Tort law has long grappled with how to deal with private information that is made public without

147. To date, there has been no copyright litigation over use of the Enron emails as training data.

148. *See* J.S. Hardin et al., *Network Analysis with the Enron Email Corpus*, 23 J. STAT. ED. 2 (2015). MIT engineering professor Leslie Kaelbling purchased the raw emails files from a government contractor for $10,000 and, after significant tinkering, re-released the emails in a machine-readable format. Leber, *supra* note 145.

149. *See Machine Learning and Law*, *supra* note 44, at 90–93 (explaining how AI learns to filter spam by analyzing massive datasets).

150. *See generally* BETHANY MCLEAN & PETER ELKIND, THE SMARTEST GUYS IN THE ROOM (2003).

151. *See, e.g.*, SAIF MOHAMMAD & TONY YANG, arXiv:1309.6347, TRACKING SENTIMENT IN MAIL: HOW GENDERS DIFFER ON EMOTIONAL AXES (2013), https://arxiv.org/abs/1309.6347 [https://perma.cc/SA7H-LAS6] (examining gender differences in how men and women use language in the Enron emails); Vinodkumar Prabhakaran et al., *Gender and Power: How Gender and Gender Environment Affect Manifestations of Power*, *in* THE 2014 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING: PROCEEDINGS OF THE CONFERENCE 1965 (2014) (demonstrating how gender information in the Enron email corpus can be used to predict the direction of hierarchical power between email participants).

152. Leber, *supra* note 145.

153. Julie Cohen, *The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy*, PHIL. & TECH. 1, 1 (2017); *see also* Karen Levy, *Intimate Surveillance*, 53 IDAHO L. REV. 679 (2016) (detailing how public and semi-public information can become fodder for surveillance).

consent.[154] In recent years, copyright law has also struggled with distinguishing private from public, as I have previously examined in the context of "revenge porn."[155] Yet we are only beginning to grapple with the tension between what is legal and what is ethical when it comes to training AI systems.[156] The appropriation of online profiles and hacked emails illustrate the ethical questions raised by treating these works as BLFD.

In 2016, for example, a researcher publicly released a dataset comprised of information scraped from the profiles of 70,000 users of the online dating site OKCupid.[157] Later that same year, WikiLeaks published more than 20,000 hacked personal emails from Hillary Clinton campaign manager, and former White House chief-of-staff, John Podesta.[158] WikiLeaks made each Podesta email available in a searchable, machine-readable format and included the raw emails in a

154. *See* William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383 (1960); Amanda Levendowski, Note, *Using Copyright to Combat Revenge Porn*, 3 N.Y.U. J. INTELL. PROP. & ENT. L. 422, 434 (2014) (discussing *Wood v. Hustler Magazine, Inc.*, 736 F.2d 1084 (5th Cir. 1984) and *Douglass v. Hustler Magazine, Inc.*, 769 F.2d 1128 (7th Cir. 1985), both featuring false light privacy tort claims in connection with the nonconsensual disclosure of intimate images).

155. *See generally* Levendowski, *supra* note 154 (analyzing how tort and copyright law have treated the disclosure of nude images without consent of the pictured individual, a phenomenon better known as "revenge porn"); *see also, e.g.*, Joseph Cox, *70,000 OKCupid Users Just Had Their Data Published*, VICE: MOTHERBOARD (May 12, 2016, 10:44 AM), https://motherboard.vice.com/en_us/article/8q88nx/70000-okcupid-users-just-had-their-data-published [https://perma.cc/GGG8-T2V3] (interviewing computer-science academics who concurred that the OKCupid data scraping was unethical).

156. Simply because works are publicly available online does not necessarily mean that using those works to train AI systems is an easy ethical question. *See infra* notes 157 (using personal data scraped from dating website for research), and 160 (using hacked emails from John Podesta to train AI email-management system). A researcher who trained a facial-recognition AI system using YouTube videos of people documenting their transition, for example, was criticized for including those people without their consent. *See* James Vincent, *Transgender YouTubers Had Their Videos Grabbed to Train Facial Recognition Software*, VERGE (Aug. 22, 2017), https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset [https://perma.cc/6GCP-YKYU]. For an examination of the ethical challenges posed by this research, see Anna Lauren Hoffman (@annaeveryday), TWITTER (Aug. 22, 2017, 7:20 PM to 7:53 PM), https://twitter.com/annaeveryday/status/900135734748758016 [https://perma.cc/Y3T8-U6QL].

157. EMIL O. W. KIRKEGAARD & JULIUS D. BJERREKÆR, THE OKCUPID DATASET: A VERY LARGE PUBLIC DATASET OF DATING SITE USERS (2016) (the dataset has since been removed).

158. Julian Assange, *The Podesta Emails; Part One*, WIKILEAKS (Oct. 11, 2016), https://wikileaks.org/podesta-emails/press-release [https://perma.cc/6WEC-BDF4].

downloadable dataset.[159] Both datasets are low-friction for training AI systems—and both are fraught with ethical quandaries.[160]

Similarly, ethical questions are posed by the availability of state and government-create BLFD, much of it is tailored to extending surveillance.[161] Earlier this year, the Department of Homeland Security announced that it would be partnering with Kaggle, recently acquired by Google, to host a competition to improve "threat-recognition algorithms" for the agency.[162] The Transport Security Administration (TSA) published its dataset for competitors, but none of the examples are real passengers. TSA claimed that it created staged images to "protect privacy," but its approach ensures that the agency has authorization to use the images freely—and incentivizes commercial AI creators to focus on systems that can use TSA-created data.[163]

The biases encoded in BLFD, like the Enron emails, are picked up by AI systems trained using those data.[164] As discussed below, the biases

---

159. *See* Email from John Podesta to Peter Huffman, Re: Risotto (Sept. 9, 2015, 2:50 AM), https://wikileaks.org/podesta-emails/emailid/4723 [https://perma.cc/QSB7-R9DZ].

160. The Podesta emails have already been used to train commercial AI systems, like Zero's AI-powered email assistant, which was referred to as "Hillary" within the company. *See* Alexander Volkov, *How We Used Hillary Clinton's Emails to Train Our AI Engine and Why You Might Want to Use Zero Email App*, ZERO APP (Sept. 8, 2016), http://zeroapp.email/blog/2016/09/08/how-we-used-hillary-clintons-emails-to-train-our-ai-engine-and-why-you-might-want-to-use-zero-email-app/ [https://perma.cc/L6JU-ERRN].

161. Works created by the United States government, for example, are not protected by copyright law. 17 U.S.C. § 105 (2012). Many states and municipalities have adopted similar provisions exempting works created by government employees from copyright protection.

162. Department of Homeland Security, *Passenger Screening Algorithm Challenge: Improve the Accuracy of the Department of Homeland Security's Threat Recognition Algorithms*, KAGGLE (June 22, 2017), https://www.kaggle.com/c/passenger-screening-algorithm-challenge [https://perma.cc/N9M5-KDVS].

163. John Mannes, *The Kaggle Data Science Community Is Competing to Improve Airport Security with AI*, TECHCRUNCH (June 22, 2017), https://techcrunch.com/2017/06/22/the-kaggle-data-science-community-is-competing-to-improve-airport-security-with-ai/ [https://perma.cc/UDL7-4YQD]. There is also an important ethical critique of diversifying datasets by targeting works created by or featuring marginalized individuals, given both the homogeneity of creators and disparate government focus on racial minorities. For a compelling discussion of this ethical quandary, see Hassein, *supra* note 30; Nabil Hassein (@NabilHassein), TWITTER (Aug. 15, 2017, 10:50 PM), https://twitter.com/NabilHassein/status/897651737296855040 [https://perma.cc/3ZES-RQKD] ("But considering who mostly controls facial recognition software, I argue Black folks won't benefit from it getting better at recognizing us."); Simone Browne (@wewatchwatchers), TWITTER (Feb. 10, 2018, 11:36 AM) https://twitter.com/wewatchwatchers/status/962364556218597376 [https://perma.cc/3UEX-K6BP] ("We need a politics of refusal around facial recognition technology (its use to reinforce borders, to criminalize, in capturing 'insurgents,' face reading drones), but that ship has almost completely sailed.").

164. Aylin Caliskan et al., *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases*, 356 SCI. 183 (2017), http://science.sciencemag.org/content/356/6334/183 [https://perma.cc/FXS3-SFAP].

embedded in two appealing, less ethically fraught sources of BLFD—public domain works and Creative Commons-licensed works—can similarly result in biased AI systems.

1. Public Domain Works

Public domain works include works that are no longer protected by copyright law. Copyright protection extends to "original works of authorship fixed in any tangible medium of expression," but that protection is neither infallible nor infinite.[165] Noncompliance with former formalities, such as proper publication and timely renewal, dedicates a work to the public domain.[166] And seventy years after an author's death, protected works likewise enter the public domain.[167] Public domain works are freely available for anyone to use, and these works can be of tremendous commercial and social value.

Publishers, like the former Penguin Books, have developed brand loyalty by investing in iconic design for public domain works.[168] New film houses relied on showing popular public domain works to develop an audience.[169] And online repositories, like Project Gutenberg and the internet Archive, have brought the richness of the public domain to anyone with an internet connection by hosting machine-readable versions of public domain texts. Digitized public domain works are

---

165. 17 U.S.C. § 102(a).

166. Rebecca Tushnet, *Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 YALE L.J. 535, 543 (2004). *Metropolis*, the cinematic masterpiece featuring a society of robot workers, was briefly in the public domain because its copyright was not re-registered. *See* Amicus Brief of Peter Decherney in Support of Petitioners, Golan v. Holder, 565 U.S. 302 (2012) (No. 10-545) [hereinafter Amicus Brief of Peter Decherney].

167. 17 U.S.C. § 302(a) (copyright protection terminates seventy years after the death of the author). Works with forfeited copyrights are also examples of works that have fallen into the public domain. *Id.* § 405 (forfeiture possible until March 1, 1989). Authors may also dedicate their works to the public domain using a license, such as Creative Commons' CC0 license. *See CC0—"No Rights Reserved,"* CREATIVE COMMONS, https://creativecommons.org/share-your-work/public-domain/cc0/ [https://perma.cc/3VPX-VY42].

168. *See* Bill Goldstein, *Publishers Give Classics a Makeover*, N.Y. TIMES (Feb. 10, 2003), http://www.nytimes.com/2003/02/10/business/media-publishers-give-classics-a-makeover.html [https://perma.cc/M9HR-HN6R]; Edwin McDowell, *Publishing: Nickleby is S.R.O.*, N.Y. TIMES (Oct. 23, 1981), http://www.nytimes.com/1981/10/23/books/publishing-nickleby-book-is-sro.html [https://perma.cc/VPD5-AD87]. Indeed, the design of Penguin Classics book jackets are so iconic that Penguin Random House released a series of postcards based on the covers. *Postcards from Penguin*, PENGUIN RANDOM HOUSE, http://www.penguinrandomhouse.com/books/307403/postcards-from-penguin-by-penguin/9780141044668/ [https://perma.cc/65Q9-EKNP].

169. *See* Amicus Brief of Peter Decherney, *supra* note 166. ("[S]ome of the most popular public domain works, like *Metropolis* (1927), *The Third Man* (1949), and Alfred Hitchcock's British films are no longer available to new distributors because their copyrights have been restored.").

easily accessible and unrestricted by copyright law, which makes these works well-suited for training AI systems specializing in NLP.[170]

Most public domain works were published prior to 1923, back when the "literary canon" was wealthier, whiter, and more Western than it is today.[171] A dataset composed exclusively of these works would exclude voices that were never recorded or rarely published, such as those of black, women, and LGBTQ authors, let alone an author who identified as all three. A dataset reliant on works published before 1923 would reflect the biases of that time, as would any AI system trained with using that dataset.

Consider the word "queer." In 1894, John Douglas, the 9th Marquess of Queensberry, transitioned the meaning of the word "queer" from an adjective meaning strange or odd to a slur for gay men when he used the term "Snob Queers" after discovering his son was romantically involved with playwright and poet Oscar Wilde.[172] However, an AI system reliant on works published prior to 1923 would never be exposed to the reclamation of the word "queer" as a term of empowerment.[173] Thus, an AI personal assistant trained exclusively using data derived from works published before 1923 would be incapable of offering adequate

---

170. *Supra* text accompanying notes 132–33("NLP techniques can help AI creators streamline and improve language-centered AI systems, from translation and text prediction to search results and conversational chatbots. But to do so effectively, NLP algorithms rely on tremendous amounts of human-generated data to 'learn' how humans use the written word.")

171. 17 U.S.C. § 304(b). The term of copyright protection around the globe, often at the United States' behest, has trended toward longer terms with fewer formalities. *Compare* Copyright Act of 1790, ch. 15, 1 Stat. 124 (repealed 1831) (granting authors a fourteen-year term with a single fourteen-year renewal), *with* Sonny Bono Copyright Term Extension Act (CTEA) of 1998, Pub. L. No. 105-298, 112 Stat. 2827 (1998) (codified as amended at 17 U.S.C. §§ 301–04) (extending the term to life of the author, plus seventy years, requiring no renewal). In the United Sates, no new works will enter the public domain until January 1, 2019. *Public Domain Day: January 1, 2018*, DUKE UNIV. SCH. OF LAW: CTR. FOR STUDY OF PUBLIC DOMAIN, https://law.duke.edu/cspd/publicdomainday/ [https://perma.cc/9TZJ-2PQ2] ("When Congress changed the law, it applied the term extension retrospectively to existing works, and gave all in-copyright works published between 1923 and 1977 a term of 95 years. The result? None of those works will enter the public domain until 2019, and works from 1961, whose arrival we might otherwise be expecting January 1, 2018, will not enter the public domain until 2057."). For a historical account of copyright term extensions, see Pamela Brannon, *Reforming Copyright to Foster Innovation: Providing Access to Orphan Works*, 14 J. INTELL. PROP. L. 145, 152–53 (2006).

172. Letter from John Queensberry, Marquess of Queensberry, to Alfred Montgomery, private secretary to the Marquess of Wellesley (Nov. 1, 1894), *quoted in* ASHLEY H. ROBINS, OSCAR WILDE–THE GREAT DRAMA OF HIS LIFE: HOW HIS TRAGEDY REFLECTED HIS PERSONALITY (2011); *see also Queer*, OXFORD ENGLISH DICTIONARY (2017), https://en.oxforddictionaries.com/definition/queer [https://perma.cc/2BVV-D4HB].

173. ERIN J. RAND, RECLAIMING QUEER (2014).

responses to an LGBTQ teen who asks whether she might be queer.[174] Such a system would even fail to recognize the acronym "LGBTQ."[175]

The voices reflected in published works have diversified since 1923, and so has language itself. Using public domain works as training data for AI systems is understandably appealing: massive numbers of public domain works are easily available in machine-readable formats online, and these works are less than legally low-risk—public domain works pose no legal risk of copyright infringement. This flavor of BLFD can easily erase entire perspectives and replicate the biases of a more homogenous authorship and less tolerant society.

2. Creative Commons-Licensed Works

Creative Commons began in 2001 as a modest nonprofit dedicated to helping creators legally share knowledge and make their works more accessible.[176] Shortly after, Creative Commons introduced its eponymous licenses, which empower creators to license their "works

---

174. However, access is no guarantee of representation. Despite access to training data reflecting modern understandings of women's health, AI personal assistants were initially incapable of answering even basic questions about menstruation, yeast infections, and sexual assault. *See* Soraya Chemaly, *The Problem with a Technology Revolution Designed Primarily for Men*, QUARTZ (Mar. 16, 2016), https://qz.com/640302/why-is-so-much-of-our-new-technology-designed-primarily-for-men/ [https://perma.cc/3CHL-JGYQ]; Rose Eveleth (@roseveleth), TWITTER (Mar. 18, 2017, 3:42 PM), https://twitter.com/roseveleth/status/843185838691405824 [https://perma.cc/N7KA-VA4U] ("Siri understands 'I had a heart attack,' but not 'I've been raped.'").

175. The acronym "LGB" originated in Usenet groups during the 1990s. *LGB*, OXFORD ENGLISH DICTIONARY (2017), https://en.oxforddictionaries.com/definition/lgb [https://perma.cc/Y7TL-4K2J]. The T, which stands for "trans" or "transgender," as well as the Q for "queer" or "questioning," were not adopted until even later. A related issue is observable for racial bias. Works published prior to 1923 did not use the term African American—the term was first used in 1782, but it was not widely popularized until the 1980s. Ben L. Martin, *From Negro to Black to African American: The Power of Names and Naming*, 106 POL. SCI. Q. 83 (1991); Fred Shapiro, *The Origin of "African American,"* YALE ALUMNI MAG. (Jan./Feb. 2016), https://yalealumnimagazine.com/articles/4216-the-origin-of-african-american [https://perma.cc/NEQ7-Q4YW]. A similar linguistic shift is observable in the rejection of colonial terms like "Oriental" to refer to Asians, Asian Pacific Islanders, and Asian-Americans or "Eskimo" to refer to Native Alaskans. Indeed, in 2016, President Barack Obama signed a law eliminating the words Negro, Oriental, and Eskimo from federal legislation, to be replaced with alternatives. Pub. L. No. 114-157, 130 Stat. 393 (2016). These examples are decidedly not intersectional, another concept that would not be taught to AI systems trained using works published prior to 1923: feminist scholar Kimberlé Crenshaw did not coin the term "intersectional" until 1989. Kimberlé Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 140 U. CHI. L. F. 139 (1989), https://philpapers.org/rec/CREDTI [https://perma.cc/MZ4T-P96J].

176. *See History*, CREATIVE COMMONS, https://creativecommons.org/about/history [https://perma.cc/Z7FL-A72S]; *What We Do: What Is Creative Commons*, CREATIVE COMMONS, https://creativecommons.org/about/ [https://perma.cc/2EEA-NRUR].

freely for certain uses, on certain conditions."[177] Rather than strictly withholding all of the exclusive rights granted to copyright owners under the law, CC licenses allow creators to give certain rights to the public, such as the right to freely reproduce and build upon the original work.[178]

The idea caught on. One year after launching its licenses, more than 1 million websites featured CC-licensed content.[179] Governments, intergovernmental organizations such as the United Nations and World Bank, numerous GLAM institutions,[180] and many more creators have adopted CC-licenses.[181] Now, there are more than 900 million CC-licensed objects, from traditional works like texts and videos to novel ones like 3D-printable objects, available online.[182] Creative Commons also designed its licenses so that it would be easy to "search for, discover, and use" CC-licensed works.[183] The Creative Commons mission is perhaps best realized by the Wikimedia Foundation projects.

The Wikimedia Foundation's collective projects, including Wikipedia, comprise the largest repository of CC-licensed multilingual, multimedia works.[184] There are 295 language editions of Wikipedia, and more than 31 million registered users who contribute to the encyclopedia anyone can edit.[185] The English-language version of Wikipedia alone has

177. *History*, CREATIVE COMMONS, *supra* note 176.

178. *Frequently Asked Questions*, CREATIVE COMMONS, https://creativecommons.org/faq/#is-creative-commons-against-copyright [https://perma.cc/X6ZA-HRBB]; *see also About the Licenses*, CREATIVE COMMONS, https://creativecommons.org/licenses/ [https://perma.cc/5ADL-ML9A].

179. Michael W. Carrol, *Creative Commons as Conversational Copyright*, *in* 1 INTELLECTUAL PROPERTY AND INFORMATION WEALTH: ISSUES AND PRACTICES IN THE DIGITAL AGE 445, 450 (Peter K. Yu ed., 2007).

180. GLAM is an acronym for "galleries, libraries, archives, and museums."

181. *See Government Use of Creative Commons*, CREATIVE COMMONS, https://wiki.creativecommons.org/wiki/Government_use_of_Creative_Commons [https://perma.cc/X6J6-7M58]; *GLAM*, CREATIVE COMMONS, https://wiki.creativecommons.org/wiki/GLAM [https://perma.cc/WT4M-ULLZ].

182. *State of the Commons 2016*, CREATIVE COMMONS, https://stateof.creativecommons.org/ [https://perma.cc/Y6ZC-LK5Q].

183. *Frequently Asked Questions*, CREATIVE COMMONS, *supra* note 178.

184. The most well-recognized Wikimedia Foundation projects are the Wikimedia Commons, which hosts image and video content, and Wikipedia. *See List of Major Creative Commons Licensed Works*, WIKIPEDIA, https://en.wikipedia.org/wiki/List_of_major_Creative_Commons_licensed_works [https://perma.cc/SXM7-Z88R]. Wikipedia, as a standalone project, is the second-largest repository of CC-licensed works after image and video-hosting service Flickr. *Id.* Personally, I have been an active editor of Wikipedia since 2011.

185. *Statistics*, WIKIPEDIA, https://en.wikipedia.org/wiki/Special:Statistics [https://perma.cc/5SZY-NAAW].

more than 5.5 million CC-licensed articles.[186] English Wikipedia is the fifth most-visited website on the internet.[187] Because Wikipedia is easily discoverable, fully machine readable, and CC licensed, its articles are especially appealing as training data for AI systems. Indeed, it is no wonder that Executive Director Katherine Maher recently noted that nearly every modern AI system relies on Wikipedia as the source of training data for facts.[188]

Wikipedia also has a significant gender imbalance: in 2011, only 8.5% of Wikipedia editors were women.[189] The editorship gender gap has measurable effects on the content of Wikipedia articles. For example, the language used to characterize women, as well as the meta-data and network structure of articles, marginalize women.[190] Biographical articles about women are likely to be missing important information when compared to articles about men.[191] And while AI

---

186. *Wikipedia: Size Comparisons*, WIKIPEDIA, https://en.wikipedia.org/wiki/Wikipedia:Size_comparisons [https://perma.cc/4AS6-MJP8].

187. *The Top 500 Sites on the Web*, ALEXA, http://www.alexa.com/topsites [https://perma.cc/M3MN-4D8Q]. Wikipedia is the only website hosted by a nonprofit organization ranked in the top twenty websites, per Alexa's rankings.

188. Dario Tarborelli (@ReaderMeter), TWITTER (June 9, 2017 at 4:15 AM), https://twitter.com/ReaderMeter/status/873106094528151552 [https://perma.cc/689T-3EUJ] ("Pretty much every AI, search, linked data platform of the planet gets its facts from @Wikipedia 1/2 –@krmaher #AIforGood") (embedding image of Linking Open Data diagram); *see also* Andrejs Abele & John McCrae, THE LINKING OPEN DATA CLOUD, http://lod-cloud.net/ [https://perma.cc/89Y3-K4GQ] (visualizing systems that rely on Wikipedia and other forms of open data).

189. This statistic comes from a comprehensive survey of Wikipedia editors conducted by the Wikimedia Foundation. WIKIMEDIA FOUND., WIKIPEDIA EDITORS STUDY: RESULTS FROM THE EDITOR SURVEY 3 (2011), https://upload.wikimedia.org/wikipedia/commons/7/76/Editor_Survey_Report_-_April_2011.pdf [https://perma.cc/6WAG-QXG9]. Some studies have suggested that this number is far too low. *See, e.g.*, Benjamin Mako Hill & Aaron Shaw, *The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation*, 8 PLOS ONE 1, 4 (2013), https://doi.org/10.1371/journal.pone.0065782 [https://perma.cc/763Z-AUG6] (suggesting that 22.7% of adult Wikipedia editors are female and that the total proportion of female editors was 16.1%, both higher proportions than estimations released by the Wikimedia Foundation). Other scholarship has attempted to diagnose the root cause of Wikipedia's gender gap. *See, e.g.*, BENJAMIN COLLIER & JULIA BEAR, CONFLICT, CONFIDENCE, OR CRITICISM: AN EMPIRICAL EXAMINATION OF THE GENDER GAP IN WIKIPEDIA, PROCEEDINGS OF THE ACM 2012 CONF. ON COMPUTER SUPPORTED COOPERATIVE WORK 383 (2012), https://dl.acm.org/citation.cfm?id=2145204.2145265 [https://perma.cc/934L-VSKQ] (examining possible hypotheses regarding the gender gap among Wikipedia editorship).

190. EDUARDO GRAELLS-GARRIDO & MOUNIA LALMAS, ET AL., FIRST WOMEN, SECOND SEX: GENDER BIAS IN WIKIPEDIA, PROCEEDINGS OF THE 26TH ACM CONFERENCE ON HYPERTEXT AND SOCIAL MEDIA 164 (2015).

191. Joseph Reagle & Lauren Rhue, *Gender Bias in Wikipedia and Britannica*, 5 INT'L J. OF COMM. 1138, 1138 (2011), http://ijoc.org/index.php/ijoc/article/viewFile/777/631 [https://perma.cc/5TQY-Q8U2] (observing that, relative to Encyclopædia Britannica, Wikipedia has better coverage and longer articles about women subjects, but articles about women were more likely to be missing

creators may only be using Wikipedia articles for facts, the gender gap can affect which "facts" AI systems can learn. The English Wikipedia article about New England Patriots tight-end Rob Gronkowski is nearly 4,000 words long and boasts 66 citations.[192] By comparison, Stanleyetta Titus, a revolutionary suffragette and the first woman admitted to the New York state bar, does not even have an article.[193] Even as the Wikimedia Foundation and its community of volunteer editors have prioritized diverse editorship and closing the gender gap, women remain in the minority of Wikipedia editors.[194]

Wikipedia is a critically important, socially valuable CC-licensed project. But its appeal as an exclusive, or even primary, source of training data for AI systems risks positioning Wikipedia as this generation's Enron corpus: new BLFD, same as the old BLFD.

III. INVOKING FAIR USE TO CREATE FAIRER AI SYSTEMS

Copyright law is no stranger to balancing questions of competition and access with notions of fairness. Indeed, the fair use doctrine has been used to address similar tensions among these normative concerns for literally hundreds of years. Back in 1841, Justice Story

---

than articles about men). When I edited the Wikipedia article about Barbara Ringer, one of the lead architects of the Copyright Act of 1976 and the first woman Register of Copyrights, her sex was mentioned in the first sentence of the article, and the information box noted that she was unmarried and childless. *Compare* Waacstats, *Barbara Ringer*, WIKIPEDIA (Oct. 16, 2013, 6:41 AM), *with* Levendowski, *Barbara Ringer*, WIKIPEDIA (June 5, 2014, 15:44 PM), https://en.wikipedia.org/wiki/Barbara_Ringer [https://perma.cc/2H7E-FJ5Z]. The phenomenon of "revenge porn," something that affects thousands of Americans, most of whom are women, did not have an article until I created it in 2013. *See* Levendowski, *Revenge Porn*, WIKIPEDIA (Oct. 13, 2013), https://en.wikipedia.org/w/index.php?title=Revenge_porn&oldid=576353319 [https://perma.cc/TQ2F-Y7ZX]; AMANDA LENHART ET AL., DATA & SOC'Y RESEARCH INST., NONCONSENUAL IMAGE SHARING: ONE IN 25 AMERICANS HAS BEEN THE VICTIM OF "REVENGE PORN" (Dec. 13, 2016).

192. *Rob Gronkowski*, WIKIPEDIA, https://en.wikipedia.org/w/index.php?title=Rob_Gronkowski&oldid=794259720 [https://perma.cc/D23G-FPAT].

193. Titus was an alumna of New York University School of Law, the first school in the city to admit women. *See Miss Titus to Become a Wife; Will Be Married to E.S. Werner by Mayor Strong To-day*, N.Y. TIMES, June 3, 1896, at 4, PROQUEST HISTORICAL NEWSPAPERS: THE N.Y. TIMES (1857–1922).

194. *See 2015-2016 Annual Plan*, WIKIMEDIA FOUND., https://wikimediafoundation.org/wiki/2015-2016_Annual_Plan [https://perma.cc/Q3G2-6D7X] (identifying "[p]rovid[ing] public support and strong stances on contentious issues like gender gap" as a mitigation technique for homogenous editorship resulting in lower-quality content). The annual Wikipedia Art+Feminism Edit-a-Thon was also launched to encourage more women to edit Wikipedia, as well as create and expand articles about notable women. *Art+Feminism*, WIKIPEDIA (Aug. 6, 2017, 8:48 PM), https://en.wikipedia.org/w/index.php?title=Wikipedia:Meetup/ArtAndFeminism&oldid=794247838 [https://perma.cc/L89Q-PZJF].

acknowledged that not all literal copying constituted infringement.[195] Justice Story unified the existing historical approaches to conclude that whether copying constituted infringement depended upon an inquiry into "the nature and objects of the selections made, the quantity and value of the materials used, and the degree in which the use may prejudice the sale, or diminish the profits, or supersede the objects, of the original work."[196] In other words, Justice Story planted the seedlings of the fair use doctrine.

Fair use persisted as judicially-created doctrine until 1976, when fair use was codified in the Copyright Act.[197] Under the Act, four factors rooted in Justice Story's articulation of fair use are of particular relevance when asking whether a secondary use is fair:

> (1) The purpose and character of the use;
>
> (2) The nature of the copyrighted work;
>
> (3) The amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
>
> (4) The effect of the use upon the potential market for or value of the copyrighted work.[198]

The fair use doctrine offers a flexible mechanism by which to balance the interests of copyright owners against the interests of subsequent creators and competitors, as well as the interests of the public.[199] The four factors, as Justice Story's early conception suggested, are not meant to be "treated in isolation, one from another. All are to be explored, and the results weighed together, in light of the purposes of copyright."[200] And, under the Act, a "fair use of a copyrighted work . . . is not an infringement of copyright."[201]

---

195. Folsom v. Marsh, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841).

196. *Id.*

197. 17 U.S.C. § 107 (2012); *see also* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 575 (1994) (tracing the evolution of fair use back to infringement cases brought in England under the Statute of Anne, the first copyright law enacted in 1710).

198. 17 U.S.C. § 107. Indeed, Congress intended for the Copyright Act to "restate the present judicial doctrine of fair use, not to change, narrow, or enlarge it in any way." H.R. Rep. No. 94-1476, at 66 (1976); S. Rep. No. 94-473, at 62 (1976). None of the fair use factors deal with how innovative computational technologies ought to be addressed by courts. *See* Edward Lee, *Technological Fair Use*, 83 S. CAL. L. REV. 797 (2010).

199. *See generally* Pamela Samuelson, *Unbundling Fair Uses*, 77 FORDHAM L. REV. 2537, 2602 (2009); *see also* H.R. Rep. No. 94-1476, at 66 (1976) ("The [Copyright Act] endorses the purpose and general scope of the judicial doctrine of fair use, but there is no disposition to freeze the doctrine in the statute, *especially during a period of rapid technological change*." (emphasis added)).

200. *Campbell*, 510 U.S. at 578.

201. 17 U.S.C. § 107.

If we hope to create less biased commercial AI systems, using copyright-protected works as AI training data will be key. Let's return to word2vec briefly. Even if would-be AI creators rely primarily on BLFD, like the Google News corpus, including data derived from additional copyrighted-protected works increases the overall size of the dataset, which can reduce the relative importance of BLFD. Second, as researchers and companies become aware of bias in AI systems, like the gender bias reflected in word2vec, Google could selectively supplement its dataset with additional copyrighted works suited to balance out blind spots. Third, recent studies have suggested that, at least in the context of training some AI systems, bigger is measurably better: AI systems trained using larger datasets perform more accurately than those trained with smaller ones.[202] And finally, as discussed previously, the sheer quantity of data needed for AI training—such as the number of photographs necessary to train a facial detection algorithm—is simply too large for many would-be AI creators to obtain without relying on the copyrighted works of others.[203] Indeed, Google would not have been able to create the corpus used to train word2vec, flaws and all, without relying on the copyrighted works of others.

This is not to say copyrighted works, as evidenced by the biases in the Google News corpus and Creative Commons-licensed works, are neutral or biased-free—they are not.[204] But there are significant benefits to using copyrighted works as AI training data. This Part examines how courts have traditionally balanced the competing values of access, competition, and fairness when presented with copyright questions involving innovative computational technologies. I apply each of the four statutory fair use factors to using copyrighted works to train less-biased AI systems, focusing on how courts have implicitly, and often explicitly, centered the values of public knowledge and social benefit when

---

202. *See, e.g.*, CHEN SUN ET AL., arXiv:1707.02968v1, REVISITING UNREASONABLE EFFECTIVENESS OF DATA IN DEEP LEARNING (2017), https://arxiv.org/pdf/1707.02968.pdf [https://perma.cc/74DR-Q8UX] (finding, in part, logarithmic improvement on vision tasks with increased volume of training data).

203. Facial detection, which identifies that a face is present in an image, is distinct from racial recognition, which purports to match a particular face with a specific person. As discussed, facial recognition may raise issues beyond those that copyright law is capable of handling fully. *Supra* note 23. And obtaining the copyrighted works necessary to create less biased facial detection algorithms nevertheless requires researchers, engineers, and designers to successfully both recognize bias and identify works that can be ethically datafied to mitigate that bias, which is no short order. *See supra* note 26 (citing the extensive legal and computer science scholarship examining the many sources of computational bias).

204. *See supra* Part II. Ironically, copying additional copyrighted works simply to identify which works reflect fewer biases would pose the same issues raised in section II.A.

concluding that innovative technologies constitute fair use, and I ask whether the fair use doctrine is equally capable of reaching the same conclusion in the field of AI bias.

A. Using Copyrighted Works as Training Data for AI Systems Is Highly Transformative

The first factor examines the "purpose and character" of the use, including whether the use is "of a commercial nature."[205] The central question of this inquiry, as the Supreme Court explained in *Campbell v. Acuff-Rose*,[206] is whether the use "merely 'supersede[s] the objects' of the original creation . . . or instead adds something new, with a further purpose . . . in other words, whether and to what extent the new work is 'transformative.'"[207]

Whether a secondary use is transformative has become something of a touchstone for the first factor inquiry.[208] As Judge Pierre Leval explained in his seminal article, *Toward a Fair Use Standard*, a transformative use:

> [M]ust be productive and must employ the quoted matter in a different manner or for a different purpose from the original . . . if the quoted matter is used as *raw material*, transformed in the creation of . . . new insights and understandings—this is the very type of activity that the fair use doctrine intends to protect for the enrichment of society.[209]

The use of copyrighted works as raw material, transformed as a means of creating new understandings, was at the core of the Second Circuit decision in *Authors Guild v. Google, Inc.*[210] In 2004, Google created a digital corpus of more than 20 million books by scanning those works, rendering them in a machine-readable format, and indexing them in the Google Books search engine.[211] The Google Books initiative enabled users to search for a particular term and view "just enough context surrounding the search term to help her evaluate whether the

---

205. 17 U.S.C. § 107(1).

206. 510 U.S. 569 (1994).

207. *Id.* at 579 (quoting Justice Story's discussion in Folsom v. Marsh, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841)).

208. For all its emphasis in the courts, however, the word "transformative" is not mentioned in the Copyright Act; a use may be fair without being transformative.

209. Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990) (emphasis added).

210. 804 F.3d 202 (2d Cir. 2015).

211. *Id.* at 208.

book falls within the scope of her interest."[212] The Second Circuit found that copying from an original work "for the purpose of criticism or commentary on the original or provision of information about it tends most clearly to satisfy *Campbell*'s notion of the 'transformative' purpose involved in the analysis of Factor One."[213]And although Google was undoubtedly motivated by profit, the court found that no reason why such "motivation should prevail as a reason for denying fair use over its highly convincing transformative purpose, together with the absence of significant substitutive competition, as reasons for granting fair use."[214]

In analyzing whether the use was transformative, Judge Leval also noted that a transformative use is one that "communicates something new and different from the original or expands its utility, thus serving copyright's overall objective of contributing to public knowledge."[215] Not all copies of protected works made to be used as training data for AI systems will be transformative. But creating copies of protected works as a pathway to mitigating bias in AI systems can serve as an important criticism of, and commentary on, humans' implicit biases.

There is also something highly transformative about the act of using copyrighted works as data, or what Judge Leval has called "raw material."[216] As Professor James Grimmelmann has observed, the purpose of copyrighted works is almost inextricably entangled with the idea of romantic readership, the idea that humans are "drawn to a work because something of the authors own unique humanity (as expressed in the work) resonates with their own."[217]

Not so for AI systems. As Professor Grimmelmann explains, AI systems "read" in only the most euphemistic of ways—and not out of

---

212. *Id.* at 218.

213. *Id.* at 215–16.

214. *Id.* at 219. Google Books, as a project, has been critiqued for its prioritization of technological rationality and efficiency without the values of education, service, and community reflected in librarians' approach to promoting access to information. *See* Anna Lauren Hoffman & Raina Bloom, *Digitizing Books, Obscuring Women's Work: Google Books, Librarians, and Ideologies of Access*, ADA J. GENDER, NEW MEDIA & TECH. (2016), http://adanewmedia.org/2016/05/issue9-hoffmann-and-bloom/ [https://perma.cc/H3XD-SNBL].

215. *Authors Guild*, 804 F.3d at 214.

216. Leval, *supra* note 209, at 1111.

217. James Grimmelmann, *Copyright for Literate Robots*, 101 IOWA L. REV. 657, 659 (2016). Professor Grimmelmann's article provides an excellent account of why robotic readership is fair use. For a related, but distinct, discussion of robotic authorship, see Pamela Samuelson, *Allocating Ownership Rights in Computer Generated Works*, 47 U. PITT. L. REV. 1185 (1985); Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 5 STAN. TECH. L. REV. 1 (2012).

any desire to connect with the humanity of the author.[218] Professor Matthew Sag has examined how using copyrighted works as "grist for the mill" serves a fundamentally different purpose than the one reflected in valuing works for their original expression.[219] Professor Sag characterizes such uses as "nonexpressive uses,"[220] and he is not alone in distinguishing expressive uses of copyrighted works from datafied ones. Professor Edward Lee, for example, has suggested a tripartite taxonomy—creational, operational, and output—based on how copyrighted works factor into a secondary use.[221] And Professors Maurizio Borghi and Statroula Karapapa have described similar uses that happen behind-the-scenes in the digital context, like search engine thumbnails of copyrighted images, as "non-display use[s] of digital works."[222]

Holdings that innovative computational systems are not fair use can be understood as implicitly drawing a distinction premised on expressive use.[223] Technologies that take too much but do too little, specifically

---

218. Grimmelmann, *supra* note 217, at 665.

219. Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. REV. 1607, 1608 (2009); *see also* Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007); Kelly v. Arriba Soft Corp., 336 F.3d 811, 818 (9th Cir. 2003) (use of images in a search engine was "unrelated to any aesthetic purpose").

220. Sag, *supra* note 219, at 1624.

221. Edward Lee, *Technological Fair Use*, 83 S. CAL. L. REV. 797, 843 (2010). Professor Lee notes that "it is difficult to find uses that are purely operational, where the only use of a copyrighted work is made internally within the machine." *Id.* at 843. Teaching AI, however, seems to be a quintessential example of a "purely operational" use under Professor Lee's framework.

222. Maurizio Borghi & Stavroula Karapapa, *Non-Display Uses of Copyright Works: Google Books and Beyond*, 1 QUEEN MARY J. INTELL. PROP. 21, 23 (2011). It is worth noting that the amicus brief of the Computer and Communications Industry Association (CCIA) defined both nonexpressive and operational uses collectively as "invisible uses." Brief Amicus Curae of the Comput. & Commc'n. Indus. Ass'n. in Support of Defendant-Appellant-Cross-Appellee TVEyes at 6, Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169 (2d Cir. 2015) (Nos. 15-3885(L), 15-3886(XAP)).

223. *See, e.g.*, Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 545 U.S. 913, 945 (2005) (noting that "there has been no finding of any fair use" for unauthorized peer-to-peer distribution of copyrighted digital music); A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1013 (9th Cir. 2001) (rejecting fair use for unauthorized peer-to-peer distribution of copyrighted digital music); Disney Enters., Inc. v. VidAngel, Inc., 224 F. Supp. 3d 967 (C.D. Cal. 2016); (granting preliminary injunction against video service that filtered objectionable content for customers). Notably, the Supreme Court did not address fair use in *American Broadcasting Companies v. Aereo* other than to note that the doctrine can "help prevent inappropriate or inequitable applications of the [Transmit] Clause." ___ U.S. ___, 134 S. Ct. 2498, 2511 (2014) (citing Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984)). And even in *Authors Guild, Inc. v. HathiTrust*, in which the digital library was ultimately found to be first use, the Second Circuit rejected the District Court's conclusion that using digital copies to facilitate access to the disabled was not transformative,

serving the same expressive purpose as the original, cannot be transformative.[224] In *Associated Press v. Meltwater*,[225] for example, Meltwater offered a digital media-monitoring service that scraped news articles from websites, including that of the Associated Press, and provided excerpts of these stories to its subscribers.[226] The Associated Press alleged that Meltwater's practices infringed the copyright in AP stories.[227] Judge Denise Cote rejected Meltwater's contentions that the purpose and use of its news reports and excerpts were transformative, countering that Meltwater not only marketed its services as substitutes, but that Meltwater offered "an expensive subscription service that markets itself as a news clipping service, not as a publicly available tool to improve access to content across the Internet."[228]

Professor Sag and other scholars have drawn distinctions between these uses, but I suggest that the language we use to describe how humans and AI systems experience copyrighted works reveals a new and different purpose. When humans experience these works, we call them "works." When AI systems do it, these works are transformed into "data." A best-selling novel becomes data about how humans use language; a selfie becomes data about the features of the human face; a conversation from a film becomes data about human voices.

B. AI Systems Rely on Copyrighted Works for Their Factual Nature

The much-maligned, oft-marginalized second factor requires an examination of the "nature of the copyrighted work."[229] The Supreme Court has stated that "fair use is more likely to be found in factual works than in fictional works," noting that "a use is less likely to be deemed fair when the copyrighted work is a creative product."[230] Courts often

---

characterizing the statement as a "misapprehension; providing expanded access to the print disabled is not 'transformative.'" 755 F.3d 87, 101 (2d Cir. 2014).

224. This was the theme advanced by Dale Cendali in the Harry Potter Lexicon case. Elyssa A. L. Spitzer, *Lawyer Curses Potter Copyright Crimes*, HARV. CRIMSON (Nov. 7, 2008), http://www.thecrimson.com/article/2008/11/7/lawyer-curses-potter-copyright-crimes-the/ [https://perma.cc/46HH-53E9]; *see also* Warner Bros. Entm't Inc. v. RDR Books, 575 F. Supp. 2d 513 (S.D.N.Y. 2008).

225. 931 F. Supp. 2d 537 (S.D.N.Y. 2013).

226. Assoc. Press v. Meltwater U.S. Holdings, Inc., 931 F. Supp. 2d 537, 541 (S.D.N.Y. 2013). I was a summer associate at Davis Wright Tremaine LLP when this case was pending before the Southern District of New York.

227. *Id.*

228. *Id.* at 554.

229. 17 U.S.C. § 107 (2012).

230. Stewart v. Abend, 495 U.S. 207, 237 (1990).

take this inquiry too literally by ignoring the broad, multidisciplinary question posed by the second factor to focus on the two subfactor considerations identified by Professor Barton Beebe: whether the work is creative or factual and whether it is published or unpublished.[231]

In *A.V. ex rel. Vanderhye v. iParadigms*,[232] however, the Fourth Circuit took a more nuanced approach to analyzing the second factor.[233] iParadigms developed a digital plagiarism detection service called Turnitin, which provides an automated way for high school and college teachers to confirm that students' work is original, rather than plagiarized.[234] Several students sued, alleging that Turnitin created infringing copies of their works of fiction and poetry.[235] Judge Traxler distinguished that a highly creative, and thus highly protected, work could nevertheless be used in a way that is unconcerned and uninterested in those creative aspects.[236] In much the same way that using a copyrighted work in litigation is considered fair use,[237] using creative works for factual purposes does not weigh against the defendant in a finding of fair use.[238] Indeed, Judge Traxler noted that using works "as part of a digitized database from which to compare the similarity of typewritten characters used in other student works—is likewise unrelated to any creative component."[239] In holding that Turnitin's service was a fair use, the court affirmed the district court finding that the digital antiplagiarism service "provide[d] a substantial public benefit through the network of educational institutions using Turnitin."[240]

Transforming copyrighted works into data is analogous. The use of fictional works as training data for AI systems to "learn" abstract concepts about language or images is "not related to any creative component" of the copyrighted works. It is, however, important for creative works to be used to train AI systems. Improving NLP includes exposing AI systems to turns of phrase, like analogies, euphemisms, metaphors, sarcasm, similes, as well as written vernacular that appear in

231. Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Options 1978–2005*, 156 U. PENN. L. REV. 549, 610 (2008).

232. 562 F.3d 630 (4th Cir. 2009).

233. *See* A.V. *ex rel.* Vanderhye v. iParadigms, LLC, 562 F.3d 630 (4th Cir. 2009).

234. *Id.* at 634–35.

235. *Id.* at 641.

236. *Id.* at 641–42.

237. *See* Bond v. Blum, 317 F.3d 385 (4th Cir. 2003), *cert. denied*, 540 U.S. 820 (2003).

238. *iParadigms*, 562 F.3d at 641.

239. *Id.* at 641–42.

240. *Id.* at 638.

non-fiction writing but are more abundant in fiction. Incorporating copyrighted works into a dataset calibrated to reduce the importance of BLFD likewise provides a "substantial public benefit."

C. Copying Entire Works to Train AI Systems Takes a Reasonable Amount and Substantiality of the Copyrighted Works

The third factor analyzes the "amount and substantiality of the portion [taken]."[241] Copying a work in full, however, "does not preclude fair use per se, [though] copying an entire work militates against a finding of fair use."[242] Cases involving innovative computational technologies regularly feature the wholesale copying of literary and visual works,[243] and courts have consistently held that wholesale copying can be necessary for certain purposes.

When a company called Arriba Soft created an internet search engine for small images, the Ninth Circuit acknowledged that copying each of a photographer's images "was reasonable . . . in light of Arriba's use of the images. It was necessary for Arriba to copy the entire image to allow users to recognize the image and decide whether to pursue more information about the image or the originating website."[244] As the Second Circuit observed in *Authors Guild v. Google, Inc.*, a case also involving digitizing books to create a searchable library, "unchanged copying has repeatedly been found justified as fair use when the copying was reasonably appropriate to achieve the copier's transformative purpose and was done in such a manner that it did not offer a competing substitute for the original."[245]

---

241. 17 U.S.C. § 107 (2012).

242. Worldwide Church of God v. Phila. Church of God, Inc., 227 F.3d 1110, 1118 (9th Cir. 2000).

243. *See, e.g.*, Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015) (copying entire literary works to create accessible, searchable database of books); Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014) (copying entire literary works to create searchable database of books); *iParadigms*, 562 F.3d at 640 (copying entire works of fiction and poetry to create plagiarism-detection service); Perfect 10, Inc. v. Amazon.com, 508 F.3d 1146 (9th Cir. 2007) (copying entire photographs to create thumbnail images for search engines); Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003) (copying entire photographs to create thumbnail images for search engines).

244. *Arriba Soft Corp.*, 336 F.3d at 821.

245. *Authors Guild*, 804 F.3d at 221; *see also HathiTrust*, 755 F.3d at 97–98 (justifying as transformative fair use the digital copying of copyrighted work for the purpose of permitting searchers to determine whether its text employs particular words); *iParadigms*, 562 F.3d at 638–40 (justifying as transformative fair use purpose the complete digital copying of a manuscript to determine whether the original included matter plagiarized from other works); *Perfect 10, Inc.*, 508 F.3d at 1165 (justifying as transformative fair use purpose the use of a digital, thumbnail copy of the

Creating wholesale copies of copyrighted literary and visual works is not always fair use. But courts have been especially willing to recognize copying the entirety of works as reasonable when those works are not exposed to the public—or at least, not exposed completely. In *Google, Inc.*, for example, Judge Leval highlighted that "Google makes an unauthorized digital copy of the entire book, [but] it does not reveal that digital copy to the public."[246] And in *HathiTrust*, search results returned page numbers on which the term appeared but "[did] not display to the user any text from the underlying copyrighted work."[247] In its recent decision in *TVEyes*, however, the Second Circuit determined that the TVEyes Watch function, which allowed subscribers to view up to ten-minute clips of copyrighted content, was "radically dissimilar to the service at issue in *Google Books*" because the length of the visible clips "likely provide TVEyes's users with all of the . . . programming that they seek and the entirety of the message conveyed . . . to authorized viewers of the original."[248]

This poses a tricky tautology for AI training data: unlike in *Google Books* and *TVEyes*, the data used to train AI systems is rarely released publicly. While this lack of transparency may strengthen companies' arguments that using copyrighted works as AI training data ought to be fair use, the ability to audit and augment biased training data *depends* on the public availability of at least some portion of that training data.[249] There remains a strong argument that, even if that data were released publicly, it would not be released for others to perceive the expressive qualities of those works but rather to identify bias in AI systems, and potentially teach other AI systems to be less biased.[250] But the Second Circuit's proclamation in *TVEyes* that less is more, at least when it comes to how much of a copy is viewable to the public, creates ample incentives for commercial AI creators to keep their training data inaccessible.

---

original to provide an internet pathway to the original); *Arriba Soft Corp.*, 336 F.3d at 818–19 (same).

246. *Authors Guild*, 804 F.3d at 221.

247. *HathiTrust*, 755 F.3d at 91.

248. Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169, 179 (2d Cir. 2018).

249. *See supra* Part II.

250. *Cf. TVEyes, Inc.*, 883 F.3d at 179 (distinguishing *Google Books* from *TVEyes*, noting that Google's snippets made it "nearly impossible for a user to see a meaningful exposition of what the author originally intended to convey to readers").

D. AI Training Data Does Not Harm the Commercial Market for Copyrighted Works

The fourth factor centers on "the effect of the use upon the potential market for or value of the copyrighted work."[251] In particular, courts focus on whether the secondary use "may serve as a market substitute for the original."[252] Courts have consistently rejected allegations that a transformative work can serve as a substitute for the original. Indeed, in *HathiTrust*, the Second Circuit flatly stated that "[a] transformative work . . . serves a new and different function from the original work and is not a substitute for it."[253]

As discussed previously, there is some market for licensing works for use as AI training data.[254] But the Copyright Act does not entitle copyright owners to profit maximization. As Judge Cote explained in *Meltwater*, "courts consider only the loss to potential licensing revenues from 'traditional, reasonable, or likely to be developed markets.'"[255] Adopting a rule requiring a license to use copyrighted works as AI training data would restrict access and further favor incumbents, as illustrated by the Google News example. It would also restrict competition, as such licenses would likely be exclusive to favorable licensees—the likelihood of AI creators licensing proprietary datasets to researchers, journalists, or competitors searching for bias may be even more remote.

Judge Barrington Parker put the tension between market harm and fair uses a bit more bluntly in *HathiTrust*: "[l]ost licensing revenue counts under Factor Four only when the use serves as a substitute for the original and [this use] does not."[256] Using copyrighted works as training data for AI systems is not a substitute for the original expressive use of the works.

Prior to the Second Circuit decision in *TVEyes*, copyright jurisprudence would suggest that, given the countervailing interests in preserving access and competition, the mere hypothetical existence of an exploitable market would be unlikely to preclude making fair use of copyrighted works as training data for AI systems. The fourth factor

---

251. 17 U.S.C. § 107 (2012).

252. Campbell v. Acuff-Rose Music, Inc., 510 U.S 569, 587 (1994).

253. Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 96 (2d Cir. 2014).

254. *See supra* Part II.

255. Assoc. Press v. Meltwater U.S. Holdings, Inc., 931 F. Supp. 2d 537, 560 (2013) (citing Am. Geophysical Union v. Texaco, Inc., 60 F.3d 913, 930 (2d Cir. 1994)).

256. *HathiTrust*, 755 F.3d at 100.

analysis in *TVEyes*, however, threatens to upset that longstanding approach to defining and assessing market harms. Judge Newman stated that the success of TVEyes's business model—selling subscriptions to access 10-minute clips collected from 1,400 television channels' broadcasts[257]—"demonstrates . . . that this market is worth millions of dollars in the aggregate" and concluded that there must be a "plausibly exploitable market for . . . [the] content."[258] Characterizing TVEyes's market as one plausibly exploitable by Fox overlooks that TVEyes's market is worth millions of dollars *because* of its aggregation—the market would look undoubtedly different for a service offering only content from a single network.[259] A similar conundrum is posed by relying on fair use to train AI systems using copyrighted works: while there may be a licensing market, the value of datasets is derived from a diversity of sources and content.

CONCLUSION

AI systems are biased because humans are biased. Indeed, AI systems learn to be all *too* human from reading, viewing, and listening to human-created works. This Article examined how copyright law has the power to channel AI in a fundamentally biased direction by looking at how copyright law can create or promote biased AI systems.

Copyright law has the power to bias AI systems, but copyright law also has the profound power to unbias them. The normative values embedded in the tradition of fair use align ultimately with the goal of mitigating bias. Fair use can, quite literally, promote creation of fairer AI systems. But the conversation cannot end there. Distinguishing between what is legally permissible and what is ethically acceptable remains an urgent question that demands rigorous engagement and thoughtful reflection by AI creators, policymakers—and perhaps even lawyers.

---

257. Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169, 179 (2d Cir. 2018).

258. *Id.* at 180. This appears to be the first time that the "plausibly exploitable market" test has been invoked in United States copyright law jurisprudence. Historically, the fourth factor required a more nuanced balancing the public benefit of the use against the personal gain of the copyright owner. *See* Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 613 (2d Cir. 2006) ("This analysis requires a balancing of 'the benefit the public will derive if the use is permitted and the personal gain the copyright owner will receive if the use is denied.'" (quoting MCA, Inc. v. Wilson, 677 F.2d 180, 183 (2d Cir. 1981)).

259. *See generally* Brief of Media Critics as *Amici Curiae* in Support of Defendant's Supplemental Motion for Summary Judgment, Fox News Network, LLC v. TVEyes Inc., 124 F. Supp. 3d 325 (S.D.N.Y. 2015) (No. 13-CV-5315).