4-28-2023

Equitable Ecosystem: A Two-Pronged Approach to Equity in Artificial Intelligence

Rangita de Silva de Alwis
University of Pennsylvania Carey Law School

Amani Carter
University of Pennsylvania Carey Law School

Govind Nagubandi
University of Pennsylvania Carey Law School

Michigan Technology Law Review

Article 2

2023

Equitable Ecosystem: A Two-Pronged Approach to Equity in Artificial Intelligence

Rangita de Silva de Alwis
University of Pennsylvania Law School

Amani Carter
University of Pennsylvania Law School

Govind Nagubandi
University of Pennsylvania Law School

Recommended Citation

Rangita de Silva de Alwis, Amani Carter & Govind Nagubandi, *Equitable Ecosystem: A Two-Pronged Approach to Equity in Artificial Intelligence*, 29 MICH. TECH. L. REV. 165 (2023).
Available at: https://repository.law.umich.edu/mtlr/vol29/iss2/2

EQUITABLE ECOSYSTEM: A TWO-PRONGED APPROACH TO EQUITY IN ARTIFICIAL INTELLIGENCE

Rangita de Silva de Alwis[*], *Amani Carter*[**], *and Govind Nagubandi*[***]

ABSTRACT

Lawmakers, technologists, and thought leaders are facing a once-in-a-generation opportunity to build equity into the digital infrastructure that will power our lives; we argue for a two-pronged approach to seize that opportunity. Artificial Intelligence (AI) is poised to radically transform our world, but we are already seeing evidence that theoretical concerns about potential bias are now being borne out in the market. To change this trajectory and ensure that development teams are focused explicitly on creating equitable AI, we argue that we need to shift the flow of investment dollars. Venture Capital (VC) firms have an outsized impact in determining which innovations will scale, we argue that influencing how these firms allocate the capital in their funds can ensure that issues of equity are top of mind for development teams. To shift the flow of investment dollars, we propose a two-pronged approach that will address two core drivers of the flow of investment: intellectual property (IP) and diversity. Our current IP system incentivizes a lack of transparency in the AI space frustrating attempts by third parties to assess whether AI- powered products and services are inequitable. And the current demographic makeup of VC firms and companies within the AI investment environment are out of sync with the general population, which can have negative downstream effects in terms of bias in AI. To change the existing dynamic, we argue for 1. creating a fifth category of IP for data and AI that would exchange ownership for compliance with

* Rangita de Silva de Alwis is faculty at the University of Pennsylvania Law School where she teaches a Policy Lab on AI and Bias along with other courses on international human rights. She was appointed visiting faculty at the Harvard Kennedy School of Government (2022-2023) where she served previously as the Leader- in- Practice at the Women in Public Policy Program Harvard University. She is also the Hillary Rodham Clinton Global Gender Equity Fellow at the Georgetown Institute for Women, Peace and Security and Senior Fellow at the Harvard Law School Center on the Legal Profession. This article grew out of her affiliation as a Fellow with the Harvard Business School's Private Capital Research Institute in 2021. She is an expert on the treaty body to the Convention on the Elimination of Discrimination against Women (CEDAW). She is grateful to the Hon. Nicole Ameline, the former Minister for Equity in France and current member of the CEDAW Committee for her inspiration.

** University of Pennsylvania Law School, J.D., 2022; Davidson College, B.A., 2016.

*** University of Pennsylvania Law School, LL.M, expected May 2023; B.S., 2009, Cornell University.

a human rights framework and 2. establishing a tax incentive for VC firms graded favorably on our commitment index. Our approach is designed to create an equitable ecosystem of sorts, one that both necessitates and encourages equitable AI from conception to implementation.

TABLE OF CONTENTS

Artificial Intelligence (AI) and Machine Learning (ML) are poised to radically transform our world. Every day we interact with AI-powered products and services. Each time we ask Alexa the weather, scroll through social media, or pay for a snack using facial recognition technology we are leveraging AI's vast potential to improve our daily lives. Experts predict that AI will transform entire industries, from banking and retail to farming and manufacturing, and will contribute to the growth of nascent technologies such as virtual reality, autonomous vehicles, and robotics.[1] AI will touch every aspect

1. Bernard Marr, *5 Reasons Why Artificial Intelligence Really Is Going to Change Our World,* FORBES (May 8, 2020), https://www.forbes.com/sites/bernardmarr/2020/05/08/5-reasons-why-artificial-intelligence-really-is-going-to-change-our-world/?sh=7d229fb378b6; *see also* Katherine Gammon, *5 Ways Artificial Intelligence Will Change the World by 2050*, U.S.C.

of our lives including critical aspects such as housing, employment, and healthcare.[2] Given AI's extraordinary potential, it is imperative that technologists, thought leaders, and lawmakers alike consider the ramifications of building AI without attention to equity. To see the consequences of developing and deploying AI tools without attention to equity we need to look no further than the controversy surrounding Facebook's role in housing discrimination beginning in 2016.[3]

Facebook's machine learning-powered advertising is arguably the company's crown jewel, accounting for the bulk of the company's revenue.[4] According to Facebook, the advertising program leverages ML models in two ways. The program utilizes ML models that can predict a particular person's likelihood of taking the advertiser's desired action based on the business objective the advertiser selects for their ad, like increasing visits to their website or driving purchases.[5] To make this prediction, the ML models consider the person's behavior on and off Facebook, the content of the ad, the time of day,

---

TROJAN FAM. MAG. (Winter 2017), https://news.usc.edu/trojan-family/five-ways-ai-will-change-the-world-by-2050.

2. *See* ROBERT SHIMONSKI, AI IN HEALTHCARE: HOW ARTIFICIAL INTELLIGENCE IS CHANGING IT OPERATIONS AND INFRASTRUCTURE SERVICES 27 (John Wiley & Sons 1st ed. 2021) (describing the ways in which artificial intelligence is being applied to healthcare operations); *see also Sony Bank to Begin Using AI to Automate the Preliminary Screening for Mortgage Loans (Summary)*, SONY FINANCIAL GROUP (May 7, 2018), https://www.sonyfg.co.jp/en/news_group/180507_01.html (announcing that Sony Bank would be using an AI application to automate preliminary screenings for mortgage loans starting in May 2018); J. Stewart Black & Patrick Van Esch, *AI-Enabled Recruiting: What is it and How Should a Manager Use It?*, 63 BUS. HORIZONS 215, 215 (2020) (arguing that AI-enabled recruiting tools have improved recruiting efficiency and that human resource managers ignore the technology at their own peril); ANAND S. RAO & GERARD VERWIJ, SIZING THE PRIZE: WHAT'S THE REAL VALUE OF AI FOR YOUR BUSINESS AND HOW CAN YOU CAPITALIZE? 1, 4–12 (Pricewaterhouse Cooper 2017) https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf (describing the immense economic potential for AI between now and 2030 and naming healthcare as one sector where we can expect to see the biggest gains).

3. *See* Katie Benner, Glenn Thrush & Mike Isaac, *Facebook Engages in Housing Discrimination With Its Ad Practices*, *U.S. Says*, N.Y. TIMES (Mar. 13, 2018) https://www.nytimes.com/2019/03/28/us/politics/facebook-housing-discrimination.html (discussing the lawsuit that the Department of Housing and Urban Development brought against Facebook for engaging in housing discrimination and mentioning that this lawsuit followed nearly three years of scrutiny of Facebook's ad-targeting practices that began with the 2016 investigation by ProPublica).

4. *See* Mike Issac, *Facebook's Profit Surges 101 Percent on Strong Ad Sales,* N.Y. TIMES (July 28, 2021) https://www.nytimes.com/2021/07/28/business/facebook-q2-earnings.html (noting that advertising revenue continues to be the bulk of Facebook's income); *see also* S. DIXON, STATISTA RESEARCH DEPARTMENT, META'S (FORMERLY FACEBOOK INC.) ADVERTISING REVENUE WORLDWIDE FROM 2009 TO 2021 (Feb. 4, 2022) https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/#:~:text=In%202020%2C%20about%2097.9%20percent,increase%20in%20comparison%20to%20the (showing that 97.9 percent of Facebook's global revenue was generated by advertising in 2020).

5. *Good Questions, Real Answers: How Does Facebook Use Machine Learning to Deliver Ads?*, FACEBOOK (June 11, 2020) https://www.facebook.com/business/news/good-questions-real-answers-how-does-facebook-use-machine-learning-to-deliver-ads.

and interactions between people and ads.[6] Additionally, the program utilizes ML models to predict an ad's quality.[7] These models consider the feedback of people viewing or hiding the ad, as well as assessments of low-quality attributes like too much text in the ad's image, sensationalized language, or engagement bait.[8] The goal of using ML in this context is to maximize value for its users and businesses alike.[9] In 2016, however, ProPublica published the results of an investigation showing that Facebook allowed advertisers to exclude specific "Ethnic Affinity" groups from seeing housing advertisements.[10] This is a prime example of what happens when we deploy tools without explicit attention to equity. While Facebook does not allow users to identify their race, Facebook's sophisticated ML models can utilize a range of signals to associate users with these "Ethnic Affinity" categories.[11] Setting aside the potential biases involved in how that kind of association is made, Facebook's initial decision to allow this feature to be used in the context of housing advertising made it possible for bad actors to engage in this kind of insidious discrimination.

The situation sounds horrifying in hindsight. When creating this capability, developers likely focused on all the positive applications for this kind of ML model. Had the possibility that bad actors may use the model for discriminatory ends been meaningfully raised, one can imagine that Facebook's team could have found a preemptive solution. The question then is: how can policymakers ensure that these discussions occur in development rooms across the country? As technologists are crafting the digital infrastructure that will power our lives, how can we be sure that they are giving due attention and care to issues of equity?

This article attempts to propose a solution to this crucial albeit complex question. We outline a systemic approach to equity in AI, one that would utilize both direct and indirect interventions that would necessitate and encourage these conversations at the ideation and development stages as opposed to post-deployment. Part I of this article discusses how inequity currently manifests in AI and ML. We discuss broadly some of the biases that

---

6. *Id.*

7. *Id.*

8. *Id.*

9. *Id.* ("Delivering personalized ads maximizes value for both people and businesses.").

10. Julia Angwin & Terry Parris Jr., *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA, (Oct. 28, 2016) https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race ("The ubiquitous social network not only allows advertisers to target users by their interests or background, it also gives advertisers the ability to exclude specific groups it calls 'Ethnic Affinities'" which the article states is a designation assigned by Facebook based on pages and posts users have liked or engaged with on Facebook).

11. Kathleen Chaykowski, *Facebook to Ban 'Ethnic Affinity' Targeting for Housing, Employment, Credit-Related Ads*, FORBES (Nov. 11, 2016), https://www.forbes.com/sites/kathleenchaykowski/2016/11/11/facebook-to-ban-ethnic-affinity-targeting-for-housing-employment-credit-related-ads/?sh=dca85de4442f.

are baked into AI and ML-powered algorithms, and we delve into the real-world implications of bias in AI through a case-study of AI-powered recruitment platforms. Part II of this article then explores our proposed solution. We posit that to ensure issues of equity are prioritized at the ideation and development stages, we need a systemic approach – one that utilizes both direct and indirect methods for facilitating the creation of equitable AI. We argue that instituting an intellectual property-based legal regime wherein the ownership of AI is made contingent upon conforming to a human-rights-centered framework will increase the likelihood that the AI- and ML-powered tools produced will be equitable. This approach is more direct because it targets the AI itself. We then argue for supplementing the direct approach with creating a diverse investment environment. This approach is indirect because it targets the people at the table rather than the AI itself. Part III then explores what this two-pronged approach could look like in practice.

For the purposes of this article, it is useful to provide a working definitional framework for both equity and AI. We consider equity to be a living term. Our core concept of equity is ensuring that each group and individual have the resources and opportunities they need to thrive, meaning it necessarily flexes depending on the social, historical, and political contexts in which it is used.[12] We distinguish equity from equality, which can be defined as providing each individual and group with the exact same resources without regard to need.[13] We rely upon a human rights framework to operationalize this conception of equity. Issues of equity and human rights are often overlapping.[14] We assert that the international legal structure built around human

12. Others have adopted this definition of equity as well. *See* Kris Putnam-Walkerly & Elizabeth Russel, *What the Heck Does 'Equity' Mean?*, STAN. SOC. INNOVATION REV. (Sept. 15, 2016), https://ssir.org/articles/entry/what_the_heck_does_equity_mean (discussing various definitions of equity and suggesting that it's "about each of us getting what we need to survive or succeed").

13. *See What's the Difference Between "Equality" And "Equity"?*, DICTIONARY.COM (Nov. 4, 2020), https://www.dictionary.com/e/equality-vs-equity/ (distinguishing between equity and equality noting that equality means things are the same while equity means things are fair); *see also 'Equity and Equality:' How they differ and overlap*, MERRIAM WEBSTER (2022), https://www.merriam-webster.com/words-at-play/equality-vs-equity-difference (noting that sameness or equal distribution are the principal denotations of equality, that they can be distinguished from justice, fairness, and impartiality the principal denotations of equity, and noting that sameness of treatment (equality) does not result in proportional fairness); *Equity v. Equality: What's the Difference – Examples and Definitions*, UNITED WAY (June 22, 2021), https://unitedwaynca.org/blog/equity-vs-equality/ (defining equality in terms of each individual or group of people being given the same resources and opportunities regardless of their circumstances).

14. *See* COLUMBIA LAW SCHOOL HUMAN RIGHTS INSTITUTE, GENDER EQUITY THROUGH HUMAN RIGHTS: LOCAL EFFORTS TO ADVANCE THE STATUS OF WOMEN AND GIRLS IN THE UNITED STATES 3 (2017), https://web.law.columbia.edu/sites/default/files/microsites/human-rights-institute/gender_equity_through_human_rights_for_publication.pdf (arguing that city, state, and local governments increasingly turn to human rights principles for guidance to address issues of gender equity); Courtney McDermott, Jewel D. Stafford, & Sharon Johnson, *Racial*

rights can serve as a powerful means for achieving equity in AI. Indeed, we argue that if AI is to be equitable it must, at a minimum, avoid violating internationally recognized human rights.[15] Human rights law is also useful insofar as AI is deployed to advance human rights. For purposes of this article, we anchor our understanding of human rights in the International Bill of Human Rights – consisting of the Universal Declaration of Human Rights, the International Covenant for Civil and Political Rights, and the International Covenant for Economic, Social and Cultural Rights and its two Optional Protocols – because it is robust and global.[16] As argued by Raso et. al., human rights law provides an agreed upon set of norms and a shared language and institutional infrastructure for ensuring AI improves our lives as opposed to violating our fundamental rights.[17] Thus, for the purposes of this article, we will consider human rights-respecting AI to be equitable AI.

AI can be a nebulous term and is used in scholarly literature to refer to a wide swath of technological advancements ranging in sophistication.[18] Generally, as set forth in Turner's work, AI can be broken down into two broad

---

Equity as a Human Rights Issue: Field Agency Practices and Field Instructors' Knowledge and Attitudes, 6 J. OF HUM. RIGHTS & SOC. WORK 14 (2021) (arguing for the importance of understanding racial equity from a human rights lens); *see also* Paul Braveman, *Social Conditions, Health Equity, and Human Rights*, 12 HEALTH & HUM. RIGHTS 31 (2010) (discussing the relationship between health equity and human rights); Rajat Khosala & Sofia Gruskin, *Equity Without Human Rights: a False COVID-19 Narrative?*, BMJ GLOB. HEALTH (2021), http://dx.doi.org/10.1136/bmjgh-2021-006720 (arguing that human rights are fundamental to achieving health equity for all); Leslie London, *'Issues of Equity Are Also Issues of Rights': Lessons from Experiences in Southern Africa,* 7 BMC PUB. HEALTH (2007) (arguing that human rights approaches succeed in achieving health equity when coupled with community engagement in ways that reinforce community capacity).

15. Others have argued similarly that for AI to benefit society writ large, it must conform with a human rights framework. *See* Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, DATA & SOC'Y 5 (arguing that in order for AI to benefit the common good, at the very least its design and deployment should avoid harms to fundamental human values and that international human rights provide a robust and global formulation of those values).

16. *Fact Sheet No.2 (Rev.1), The International Bill of Human Rights*, UNITED NATIONS HUM. RIGHTS, https://www.ohchr.org/Documents/Publications/FactSheet2Rev.1en.pdf.

17. Filippo A. Raso, Hanna Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, & Levin Kim, *Artificial Intelligence &. Human Rights: Opportunities & Risks,* BERKMAN KLEIN CTR. RSCH. PUBL'N NO. 2018-6 (2018), *available at* https://ssrn.com/abstract=3259344 or http://dx.doi.org/10.2139/ssrn.3259344.

18. *See* Chris Smith, Brian McGuire, Ting Huang, & Gary Yang, *The History of Artificial Intelligence*, UNIVERSITY OF WASHINGTON (2006), https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf (noting that the subject of AI is one of the most elusive subjects in Computer Science due in part to how large and nebulous the subject is and explaining that most of the breakthroughs in AI aren't noticeable to most people due in part to the subtle ways the technology is used); *see also The Short and Sweet On AI*, AI COLLECTIVE, https://www.aicollective.co/ai-introduction#:~:text=The%20term%20%22artificial%20intelligence%22%20is,not%20is%20another%20matter%20entirely (noting that the term "artificial intelligence" can be nebulous and pointing out that under some broad definitions even early stage calculators would qualify).

buckets: narrow and general.[19] Narrow AI is characterized by a system's ability to achieve a certain stipulated goal or set of goals in a manner or using techniques which qualify as intelligent and is only suited to the goal for which it was designed.[20] General AI, on the other hand, is capable of an unlimited range of goals.[21] General AI can set new goals independently, including in situations of uncertainty and vagueness and it is the kind of AI we typically see portrayed in popular culture.[22] Much of what we are currently able to develop falls into the category of Narrow AI. For example, the ML-powered advertising platform created by Facebook discussed above falls into this category as the algorithm takes in data points and makes predictions about users and ad performance. General AI is currently beyond our technological capabilities.[23]

The principal intervention of this article is to create a systemic approach to equity in AI, one that would remain effectual once our technological capabilities extend into the realm of General AI. Indeed, bias in AI is a complex issue, one that requires a robust and systemic approach. Regulatory frameworks that are predicated on how data and AI work today may not be workable as we develop ever more sophisticated AI-powered technology. Our approach is designed to create an equitable ecosystem of sorts, one that both necessitates and encourages equitable AI from conception to implementation.

I. Understanding the Problem: How AI currently falls short of equitable

Much scholarly attention has been paid to the theoretical discussion of bias in AI.[24] Scholars have identified several different types of bias that can

---

19. *See* JACOB TURNER, ROBOT RULES: REGULATING ARTIFICIAL INTELLIGENCE 6, Palgrave McMillan (2018).

20. *See id.* There are other ways of thinking about what constitutes AI. There is a vibrant discussion amongst thought leaders regarding how we should be thinking about AI. We use Turner's theoretical framework for understanding AI because of its conceptual accessibility.

21. *Id.*

22. *Id.*; *See* James Cameron, THE TERMINATOR, (Cinema '84 1984) (for examples of General AI in popular culture); *see also* I, ROBOT, (Davis Entertainment 2004).

23. *See* Byron Reese, *The Possibility of General AI*, VENTUREBEAT (June 26, 2022, 8:10 AM), https://venturebeat.com/2022/06/26/the-possibility-of-general-ai/ (explaining that general AI doesn't exist yet except in science fiction and that as yet no one knows how to create it); Sam Shead, *Computer Scientists Are Questioning Whether Alphabet's DeepMind Will Ever Make A.I. More Human-like*, CNBC (June 18, 2021, 3:43 AM), https://www.cnbc.com/2021/06/18/computer-scientists-ask-if-deepmind-can-ever-make-ai-human-like.html (discussing disagreement among the AI community as to whether reinforcement learning is a viable method for creating general AI).

24. *See, e.g.*, McKenzie Raub, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias, and Disparate Impact Liability in Hiring Practices*, 71 ARK. L. REV. 2, 529 (2018), https://heinonline.org/HOL/P?h=hein.journals/arklr71&i=549 [hereinafter *Bots, Bias and Big Data*]; Robert H. Sloan & Richard Warner, *Beyond Bias: Artificial Intelligence and Social Justice*, 24 VA. J.L. &. TECH. 1 (2020); Aram A. Gavoor & Raffi Teperdjian, *A Structural Solution*

manifest in AI systems design, including data bias[25] and programmer bias.[26] These biases ultimately lead to the recreation of existing modes of discrimination, which can be violative of the human rights established in the Universal Declaration of Human Rights,[27] while casting the veneer of fairness and impartiality. We will briefly examine each of these biases in turn.

Data is key for Narrow AI.[28] Data bias can manifest in several different ways, so for brevity, we will limit our discussion to data sampling.[29] Recall, Narrow AI is designed to achieve a particular goal like predicting a consumer's likelihood to buy a pair of shoes or whether a consumer is likely to repay a loan. In the broadest of terms, the way programmers and developers create certain kinds of Narrow AI models is by taking a vast pool of data points and using algorithms to mine that data for "interpretable patterns otherwise too subtle or complex for unaided human discernment."[30] This is all well and good unless that pool of data is partial or nonrepresentative.[31] Data collection itself "often suffers from biases that lead to the over- or underrepresentation of certain groups."[32] Datasets can suffer from the "nonrandom,

---

to Mitigating Artificial Intelligence Bias in Administrative Agencies, 89 GEO. WASH. L. REV. ARGUENDO 71 (2021); Rafal Rejmaniak, *Bias in Artificial Intelligence Systems*, 26 BIALOSTOCKIE STUDIA PRAWNICZE 25 (2021).

25. *See* Adam Zewe, *Can Machine-Learning Models Overcome Biased Datasets?*, MIT NEWS (Feb. 21, 2022), https://news.mit.edu/2022/machine-learning-biased-data-0221 (discussing machine learning models exhibiting bias due to the bias present in the datasets used to train that machine learning model).

26. *See* Kyle Wiggers, *Study Finds Diversity in Data Science Teams is Key in Reducing Algorithmic Bias*, VENTUREBEAT (Dec. 9, 2020, 1:10 PM), https://venturebeat.com/business/columbia-researchers-find-white-men-are-the-worst-at-reducing-ai-bias/ (discussing how bias of the engineers or programmers creating AI has been shown to impact the artificial intelligence models).

27. *See* LINDSEY ANDERSEN, HUMAN RIGHTS IN THE AGE OF ARTIFICIAL INTELLIGENCE, 18-30 (2018), https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf, (discussing how current AI uses violate or risk violating human rights as set out in the International Bill of Human Rights).

28. *See Bots, Bias and Big Data*, *supra* note 24, at 533 (explaining that for deep learning to function, algorithms need to be fed data).

29. *See* James Manyika, Jake Silberg & Brittany Presten, *What Do We Do About the Biases in AI?*, HARV. BUS. REV. (Oct. 25, 2019), https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai, (explaining biased data sampling as the over- or underrepresentation of groups in the training data).

30. *See Bots, Bias and Big Data*, *supra* note 24, at 533.

31. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 684 (2016) [hereinafter *Big Data's Disparate Impact*] ("Decisions that depend on conclusions drawn from incorrect, partial, or nonrepresentative data may discriminate against protected classes.").

32. Eirini Ntoutsi, Pavlos Fafalios, Ujwal Fadiraju, Vasileios Iosfidis, Wolfganf Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, Franco Turini, Symeon Papadopoulos, Emmanouil Krasanakis, Ioannis Kompatsiaris, Katharina Kinder-Kurlanda, Claudia Wagner, Fariba Karimi, Miriam Fernandez, Harith Alani, Bettina Berendt, Tina Kruegel, Christian Heinze, Klaus Broelemann, Gjergji Kasneci, Thanassis Tiropanis & Steffen Staab, *Bias in Data-Driven Artificial*

systemic omission of people who live on big data's margins, whether due to poverty, geography, or lifestyle, and whose lives are less 'datafied' than the general population's."[33] When the initial dataset over- or under-samples marginalized groups it can skew conclusions that are drawn by the algorithmic models during the data mining process.[34] Those skewed conclusions, when deployed in an AI system, may very well discriminate against these marginalized groups, replicating in AI the inequities that already plague our society.[35]

Relatedly, programmers are not bias-immune,[36] and even if provided a perfectly representative dataset, programmers' unconscious biases may leak into the AI system during the data labeling process. Data labeling is the process by which training data[37] is manually assigned labels by programmers or data miners.[38] Again, Narrow AI is designed to achieve a particular goal. Programmers are responsible for translating that goal into a question about the value of some target variable.[39] For example, if the goal of an AI system is predicting creditworthiness, then datasets containing information about consumer payment habits will need to be labeled. This process is ripe for unconscious bias because determining which kinds of data should be labeled defaulting and which kinds should not is likely subjective or arbitrary.[40] It is not obvious that missing four credit card payments should be labeled as defaulting while missing three is not.[41]

Intelligence Systems – An Introductory Survey, 10 WIREs DATA MINING & KNOWLEDGE DISCOVERY, 4 (2019), https://doi.org/10.1002/widm.1356.

33. Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55, 57 (2013).

34. *See Big Data's Disparate Impact* supra note 31, at 855.

35. *See id.* (discussing examples of how nonrepresentative datasets deployed in AI-powered applications can lead to discrimination against marginalized groups); *see also* Raso, *supra* note 17 at 7 (noting that misrepresentation in data can lead to vicious cycles that perpetuate discrimination and disadvantage and can occur with both under-representation of historically disadvantaged groups, for example, women and people of color in IT developer communities and image datasets, and with over-representation, for example, African-Americans in drug-related arrests).

36. *See generally* Francesca Gina & Katherine Coffman, *Unconscious Bias Training That Works*, HARV. BUS. REV. (Sept. 2021), https://hbr.org/2021/0 9/unconscious-bias-training-that-works (discussing unconscious bias and the kinds of training that may be helpful to combat it).

37. Think of "training data" as a large pool of examples. If you want a model to accomplish a task, such as differentiate between two different kinds of leaves, the model will need to be exposed to examples of different kinds of leaves. From that vast set of examples, the model then identifies certain patterns or associations, and when presented with a new leaf in the future uses those patterns and associations to make a determination as to what kind of leaf it is. For a more succinct definition. *See Big Data's Disparate Impact*, *supra* note 31, at 680.

38. *Id.* at 681 ("Labeling examples is the process by which the training data is manually assigned class labels.").

39. *See* McKenzie, *supra* note 24, at 533.

40. *See Big Data's Disparate Impact*, *supra* note 31, at 681 (discussing the subjectivity of data labeling and using creditworthiness as an example).

41. *See id.*

Unconscious bias can seep in even where data labeling is objective. Consider the creditworthiness example but where the data contains information on whether individuals had ever been unbanked.[42] The process of labeling data points as never unbanked or previously unbanked would be objective. It would be obvious whether a person had been unbanked in the past. If the programmer selected this kind of data as a target variable, however, they'd be relying on the presumption that whether a person was previously unbanked is a good measure of creditworthiness. This presumption, however, may be tinged with unconscious bias. People of color are more likely to be unbanked as communities of color have historically been underserved by banking institutions.[43] The historical decisions of banking institutions which resulted in inequity would then be transmuted into the resulting AI system. Automating processes based on data characterized by prior prejudice or implicit bias can result in creating a formalized rule that could systematically alter the prospects of all future credit applicants.[44]

To fully survey all the ways in which bias could manifest in AI development would be an expansive project, one which we do not endeavor to undertake here. Rather, we thought it critical to highlight that this is not merely a theoretical issue. We are already starting to see examples of these biases in the AI that we use daily.[45] Recent research suggests that AI-powered recruitment platforms may exhibit anti-Black bias and can impact the way that Black users interact with hiring platforms.[46] The data suggests that hiring platforms – such as LinkedIn, Indeed, Monster.com, and ZipRecruiter – may be targeting or focusing Black users' identities as much or more than their actual

42. *See* Diane Standaert, Naomi Camper, Dean Karlan & Kara Perez, *Unbanked: What It Means to Be Outside of the Banking System*, NPR (Apr. 5, 2021), https://www.npr.org/2021/04/05/984475870/unbanked-what-it-means-to-be-outside-of-the-banking-system (for a brief discussion of what it means to be unbanked).

43. Annie Nova & Darla Mercado, *Where You Bank Can Make a Difference for Racial Justice*, CNBC (July 4, 2020, 9:45 AM), https://www.cnbc.com/2020/07/04/upset-about-racial-injustice-where-you-bank-can-make-a-difference.html.

44. *See Big Data's Disparate Impact*, *supra* note 31, at 682.

45. *See* Starre Vartan, *Racial Bias Found in a Major Health Care Risk Algorithm*, SCI. AM. (Oct. 24, 2019), https://www.scientificamerican.com/article/racial-bias-found-in-a-major-health-care-risk-algorithm/ (discussing the health-care risk-prediction algorithm that exhibited racial bias in that it relied on a faulty metric for determining need). *See* Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (discussing their findings that a recidivism algorithm exhibited racial bias against black individuals). *See* Jefferey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (discussing the gender bias exhibited the Amazon's hiring tool).

46. *See* Amani Carter, *Unmaksing Coded Bias: Why We Need Inclusion & Equity in AI*, HARV. KENNEDY SCH. WOMEN IN PUB. POL'Y PROGRAM (Apr. 19, 2021), https://wappp.hks.harvard.edu/news/elephant-ai, [hereinafter *Unmasking Coded Bias*].

credentials.[47] Nearly forty-two percent of respondents reported feeling that the employment opportunities recommended to them on hiring platforms were mismatched with their credentials, with nearly thirty-five percent identifying such recommended jobs as below their qualifications.[48] Just about thirty-three percent of respondents reported feeling that the job opportunities recommended to them matched their qualifications. [49] According to the report, this suggests that the hiring AI used by respondents are almost as likely to underestimate Black respondents' abilities by recommending opportunities that are below respondents' qualifications as the AI is to correctly assess Black respondents' abilities and match them with opportunities matching their qualifications.[50] The research reported three main types of bias evidenced by its survey data: self-censoring bias, bias in design, and stereotype threat.[51] We will examine each in turn.

The research found evidence that Black applicants utilizing AI-powered hiring platforms expressed clear concern that indicating one's racial identity could limit professional opportunities, as well as a clear impulse to engage in self-censoring techniques to increase the likelihood of favorable outcomes.[52]

As noted by the report, racial discrimination in the applicant evaluation process remains a pervasive problem in North American labor markets.[53] Studies show resumes containing minority racial cues, such as a distinctively Black name, led to thirty to fifty percent fewer callbacks from employers than do otherwise equivalent resumes without such cues.[54] Black applicants respond by engaging in self-censoring techniques such as resume whitening – removing references to race with the hopes of boosting their shot at securing the employment opportunity – and evidence suggests that these techniques have proven successful.[55] As noted by the research, "widespread self-censoring in the Black community can result in a dearth of data used to train the AI that hiring platforms use to identify potential candidates which ultimately could result in less accurate recommendations for Black candidates."[56] Ultimately, self-censoring by Black applicants can result in predictive models

---

47. *Id.* at 7.
48. *Id*.
49. *Id*.
50. *Id.* at 7-8.
51. *Id.* at 10-11.
52. *Id.* at 12.
53. *Id.* at 11; *See also*, Sonia K. Kang, Katherine A. DeCelles, András Tilcsik & Sora Jun, *Whitened Resumes: Race and Self-Presentation in the Labor Market*, 61 ADMIN. SCIENCE Q. 469 (2016) [hereinafter *Whitened Resumes*].
54. *Unmasking Coded Bias*, *supra* note 46, at 11; *See also Whitened Resumes*, *supra* note 53, at 469.
55. *Unmasking Coded Bias*, *supra* note 46, at 11; *See also* Diana Gerdeman, *Minorities Who 'Whiten' Job Resumes Get More Interviews*, WORKING KNOWLEDGE: HARV. BUS. SCH., (May 17, 2017), https://hbswk.hbs.edu/item/minorities-who-whiten-job-resumes-get-more-interviews.
56. *Unmasking Coded Bias*, *supra* note 46, at 13.

that work less optimally for Black applicants, allowing anti-Black bias to leak into the AI system.[57]

The research also found evidence suggesting bias in design, specifically in terms of target variable determination and data bias.[58] The research found evidence that AI-powered hiring programs have incorporated legacy skills-based tests that have historically disadvantaged Black test-takers.[59] As discussed above, basing target variables upon legacy tests that have historically disadvantaged marginalized groups can transmute that inequity into the resulting AI system. Additionally, the research found evidence suggesting that AI-powered platforms may be suffering from datasets underrepresenting Black applicants' interests and potentially the areas where they live, which could lead to suboptimal results for Black applicants utilizing AI-powered platforms.[60] Furthermore, the report found some evidence that AI-powered hiring platforms may be relying upon datasets overrepresenting Black candidates for diversity, equity, and inclusion opportunities.[61] An algorithm that overrepresents some candidates can display inaccuracies that disadvantage the underrepresented group; alternatively such an algorithm might pigeonhole some groups, such as Black candidates being offered DEI positions because the algorithm's dataset is primed for that.[62]

Lastly, Black applicants may be experiencing stereotype threat while assembling application materials, engaging with hiring platforms, and taking AI-powered assessments.[63] To understand why this finding is important, one must understand how stereotype threat functions. Stereotype threat is defined as "a situational predicament in which individuals are at risk, by virtue of their actions or behaviors, of confirming negative stereotypes about their group.[64] The fear of stereotype confirmation "can hijack the cognitive systems required for optimal performance, resulting in poorer outcomes."[65] This is especially concerning from a bias perspective because data showing that Black applicants perform poorer overall than applicants that do not face stereotype threat may be reincorporated into the predictive models, thus allowing anti-Black bias to leave into AI systems.

57. *Id.*
58. *Id.* at 14.
59. *Id.* at 16.
60. *See id.* at 14-15.
61. *See id.* at 17.
62. *Id.*
63. *Id.* at 18-19.
64. *Id.* at 17-18.
65. *Id.* at 18.

II. EXPLORING SOLUTIONS: A TWO-PRONGED APPROACH TO CREATING AN EQUITABLE ECOSYSTEM

As we've noted, much ink has been spilled theorizing and exploring the ways in which we are building inequity into the digital infrastructure of tomorrow. We focus our attention on what technologists, thought leaders, and lawmakers should do about it. As Michael Kearns and Aaron Roth argue in their work, an algorithm is a precisely specified series of instructions for performing a concrete task.[66] For any given task, there are choices and trade-offs to be made as to optimization – one could make a model that optimizes for speed or minimal processing power or a host of other desirable outcomes.[67] It stands to reason, and Kearns and Roth agree, that if we can optimize for performance metrics like these then we can also explicitly optimize for social values such as privacy, fairness, or equity.[68] We do not endeavor to explain here all the ways that designers can introduce new goals such as equity into the code, our primary goal is to create an environment wherein designers are incentivized to make such introduction.[69] If the question is how can we ensure that development teams strive to build equity into the AI they build, we believe the answer lies with investment dollars.

Venture capital (VC) firms have backed some of the most high-growth and influential companies in the world.[70] Nearly half of entrepreneurial companies that graduate to the public marketplace have taken on VC funding.[71] Studies have shown that VC has a substantial impact on innovation at the industry level and appears to be three to four times more potent in stimulating innovation than a dollar of traditional corporate research and development.[72] Even beyond VC, research has shown that corporate investment relationships may impact how startups search for potential innovations.[73] If we can focus the flow of investment dollars towards equitable AI, we believe we can spark a powerful wave of innovation in this area. If investors are looking to back

---

66. MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM DESIGN 4 (2019).

67. *Id.* at 5.

68. *Id.*

69. *See id.* at 18-19 (for a more in-depth discussion of how designers can introduce new goals such as fairness and equity into the code).

70. Josh Lerner & Ramana Nanda, *Venture Capital's Role in Financial Innovation: What We Know and How Much We Still Need to Learn*, 34 J. OF ECON. PERSP. 237, 237 (2020).

71. *Id.*

72. Roberta Dessi & Nina Yin, *The Impact of Venture Capital on Innovation*, THE OXFORD HANDBOOK OF VENTURE CAP. 668, 668 (Douglas Cumming ed., 2012); *see also* Nawab Khan, Haitao Qu, Jing Qu, ChunMiao Wei & Shihao Wang, *Does Venture Capital Investment Spur Innovation? A Cross-Countries Analysis*, 11 SAGEOPEN 1 (2021).

73. *See generally* Francisco Polidoro, Jr. & Wei Yang, *Corporate Investment Relationships and the Search for Innovations: An Examination of Startups' Search Shift Towards Incumbents*, 32 ORG. SCI. 909 (2021).

equitable AI, then development rooms across the globe will be keen to create it.

We argue for a two-pronged approach to changing the flow of investment dollars in the AI space. The first prong focuses on shifting the ownership regime for data and AI, effectively requiring AI systems to conform with a human rights-centered framework. The second prong focuses on creating a diverse investment environment wherein both the VC firms investing in AI and the founding teams creating AI more closely reflect the broader US population. We examine each in turn.

A. *The Case for a Human-Rights Centered Framework for AI Ownership*

Currently, intellectual property, a driver of investment dollars, de-incentivizes transparency and is ill-equipped for encouraging equitable AI. IP has a sizeable impact on which companies receive investment dollars needed to deploy at scale. Patent quality, for example, has been shown to improve the size of investment and firm valuation.[74] Relatedly, research has shown that technology-based startups can utilize patent rights to communicate the quality of their underlying technologies to investors.[75] Research has also shown that Venture Capital dollars flow to IP-intensive industries.[76] Given IP's impact on investment decisions, it follows that IP is a powerful tool for directing the flow of investment dollars. We argue that the current IP protection landscape for AI is poorly suited for encouraging the creation of debiased and human rights-respecting AI. We posit that creating a new IP framework that (i) applies specifically to AI and secondarily to data, and (ii) employs a human rights framework could create a strong incentive for investors to fund only those companies creating debiased and human rights-respecting AI.

Ownership of AI under our current IP framework requires consideration of both how algorithms and algorithmic models are treated as well as how data and datasets are treated. Ultimately, data and datasets play an immensely large role in determining the contours of algorithmic models and AI systems, particularly for Narrow AI.

Presently, there are four main categories of IP recognized in the United States: copyright, trademark, patent, and trade secret. Each is designed to protect specific kinds of intangible assets. Copyright protects artistic or literary

74. *See generally* Shi Chen, Wei Meng, & Haitian Lu, *Patent as a Quality Signal in Entrepreneurial Finance: A Look Beneath the Surface*, 47 ASIA-PACIFIC J. OF FIN. STUD. 280 (2018).

75. *See generally* Carolin Haeussler, Dietmar Harhoff & Elisabeth Mueller, *How Patenting Informs VC Investors – The Case of Biotechnology* 43 RSCH. POL'Y 1286 (2014).

76. Mary Juetten, *Do Venture Capitalists Care About Intellectual Property?*, FORBES (Aug. 11, 2015, 10:23 AM), https://www.forbes.com/sites/maryjuetten/2015/08/11/do-venture-capitalists-care-about-intellectual-property/?sh=7ee3c1c15b87.

works of original authorship.[77] Trademark generally protects brands or source identifiers – those markers that distinguish a company's goods or services from the rest of the market.[78] At a high level, patents protect certain kinds of new inventions, and trade secrets protect information not known to the public that provide a company with a competitive edge because it's secret.[79] While copyright and patent law can protect certain discrete elements of the process by which AI systems are made, both regimes are an uneasy fit for protecting data for reasons we discuss in detail below. Because of this uneasy fit, trade secret protection is the IP regime of choice for protecting the datasets companies use to create and train their algorithmic models and is often used to protect the algorithmic models as well.[80] The overall result is a lack of transparency which runs counter to the goal of increasing equity in the AI space. We examine this dynamic in more detail below.

Patent law offers virtually no protection for datasets used to train algorithmic models and very narrow and uncertain protection for AI-powered software. Generally, new and useful processes, machines, manufactures, compositions of matter, or any new or useful improvement of the same are eligible for patent protection, but data doesn't quite fit into any of these categories.[81] Data is more akin to facts and facts, like laws of nature, are not patentable.[82] Certain kinds of software utilizing the algorithmic models created

77. *See Trademark, Patent, or Copyright*, U.S. PAT. & TRADEMARK OFFICE (Mar. 31, 2021, 12:00 PM), https://www.uspto.gov/trademarks/basics/trademark-patent-copyright (stating that the following is legally protected as copyright law "artistic, literary, or intellectually created works, such as novels, music, movies, software code, photographs, and paintings that are original and exist in a tangible medium such as paper, canvas, film or digital format).

78. *See id.* (stating that the following is legally protected by trademark law "a word, phrase, design, or combination that identifies your goods or services, distinguishes them from the goods or services of others, and indicates the source of your goods or services").

79. *See id.* (stating that the following is legally protected by patent law "technical inventions, such as chemical compositions like pharmaceutical drugs, mechanical processes like complex machine, or machine designs that are new, unique, and usable in some type of industry"); *see also Trade Secrets*, WORLD INTELL. PROP. ORG., https://www.wipo.int/tradesecrets/en/ (describing what qualifies as a trade secret).

80. *See* Meghan J. Ryan, *Secret Algorithms, IP Rights, and the Public Interest*, 21 NEV. L.J. 61, 62 (noting that where algorithm-based software is concerned, the law encourages trade secret protection over patent protection); *see also* Jordan Jaffe, Jared Newton, Patrick Curran, Anil Makhijani & Zack Flood, *The Rising Importance of Trade Secret Protection for AI-Related Intellectual Property*, QUINN EMANUEL URQHART & SULLIVAN, LLP (Apr. 24, 2021), https://www.quinnemanuel.com/media/wi2pks2s/the-rising-importance-of-trade-secret-protection-for-ai-related-intellec.pdf (noting that companies are increasingly turning to trade secret protection to protect their AI-related intellectual property).

81. *See* 35 U.S.C. § 101 (1952) ("Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.").

82. *General information concerning patents*, U.S. PAT. & TRADEMARK OFF., https://www.uspto.gov/patents/basics/general-information-patents (last accessed Sep. 28, 2022) ("Interpretations of the statute by the courts have defined the limits of the field of subject matter

and trained by a dataset, however, may be patentable under certain conditions. The Supreme Court proffered a combined Alice/Mayo framework consisting generally of two steps for evaluating whether the software is patentable.[83] In the first step, the court determines whether the claims at issue are directed to a patent-ineligible concept.[84] In the second step, the court examines the elements of the claim to determine whether it contains an "inventive concept" sufficient to "transform the claimed abstract idea into a patent-eligible application."[85] The second step requires assessing whether the application does more than just stating the abstract idea while adding the words "apply it."[86] Because of the way algorithmic models are generated, AI-powered software is very likely to satisfy step one and fail step two.

At a high level of generality, programmers develop algorithmic models by feeding vast amounts of training data into computer processes that identify patterns and relationships within the data, those relationships are then used to predict outcomes when the model is given new data. Those relationships could readily be characterized as "laws of nature," one of the three categories of patent ineligible concepts.[87]Additionally, many AI systems are designed to perform actions that humans previously performed.[88] This is precisely the dynamic at play in the Alice Corp. v. CLS Bank International case where the court found a settlement risk software to be directed at a patent ineligible concept. In that case, the computer software at issue involved a method of exchanging financial obligations between two parties using a third-party intermediary to mitigate settlement risk.[89] The court found that on their face, the claimed invention was directed toward the concept of intermediated settlement, i.e. the use of a third party to mitigate settlement risk.[90] The court went on to say that the concept of intermediated settlement is a "fundamental economic practice long prevalent in our system of commerce" and therefore

---

that can be patented, thus it has been held that the laws of nature, physical phenomena, and abstract ideas are not patentable subject matter.").

83. The framework is derived from two cases: i) *Mayo Collaborative Servs. v. Prometheus Labs*., Inc., 566 U.S. 66, 89 (2012) and ii) *Alice Corp. Pty. Ltd. v. CLS Bank Int'l*, 573 U.S. 208, 217 (2014).

84. *See Alice Corp.*, 573 at 218 ("We must first determine whether the claims at issue are directed to a patent-ineligible concept.").

85. *Id.* at 221.

86. *See id.* ("Mayo made clear that transformation into a patent-eligible application requires 'more than simply stat[ing] the [abstract idea] while adding the words 'apply it.'").

87. *See Mayo Collaborative Servs*., 566 U.S. at 71 (discussing why laws of nature, natural phenomena, and abstract ideas are not patentable and describing a diagnostic program as applying natural laws describing relationships between the concentration of the blood of certain metabolites and the likelihood of efficacy of a drug dosage).

88. *See* Shabbi S. Khan & Nikhil T. Pradhan, *How to Overcome the Two Biggest Challenges of Patenting AI Technologies*, FOLEY & LARDNER LLP: INNOVATIVE TECHNOLOGY INSIGHTS (Feb. 27, 2020), https://www.foley.com/en/insights/publications/2020/02/how-overcome-challenges-patenting-ai-technologies.

89. See *Alice Corp.*, 573 at 214.

90. *Id*. at 211.

is an "abstract idea" beyond the scope of what the patent law allows.[91] Consider, a lending algorithm developed by a bank to determine whether to provide a mortgage to an applicant. The algorithm would be directed towards the concept of creditworthiness, another fundamental economic process that has been prevalent in our system of commerce. Despite the fact that the algorithm may be more efficient than a human, the point remains that assessing creditworthiness would be considered an abstract idea. Thus, it is extremely likely that AI-powered software would be considered directed at a patent-ineligible concept. Further, courts struggle to determine whether claims have additional elements that would add an inventive concept sufficient to transform the claimed abstract idea into a patent-eligible application. AI-powered software might be able to clear the step two hurdle if it amounts to an improvement on an existing technological process, but that is challenging to parse and difficult to predict.

Similarly, copyright protection of the datasets used to create algorithmic models is thin[92] where available at all, and is only narrowly available for software whether AI-powered or otherwise.[93] It is important to note that copyrights protect the expression of ideas rather than the ideas themselves, and compilations can qualify as protected works.[94] The datasets that tech companies use to train algorithmic models used in AI-powered software could be categorized as a compilation, but the protection does not extend to the data within the set. The layer of creativity that is relevant for copyright purposes is the manner in which the data is arranged.[95] This protection might be useful if another company were to copy the entire dataset and use the data only as arranged,[96] but even this protection may be limited if portions of the

91. *Id.*

92. Copyright protection for datasets is limited to the way in which the data is compiled and does not extend to the data itself. "Factual compilations . . . may possess the requisite originality. The compilation author typically chooses which facts to include, in what order to place them, and how to arrange the collected data so that they may be used effectively by readers. These choices as to selection and arrangement, so long as they are made independently by the compiler and entail a minimal degree of creativity, are sufficiently original that Congress may protect such compilations through the copyright laws." Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 348 (1991).

93. *See* Ryan, *supra* note 80, at 73-75 ("Although the uncertainty about the applicability of copyright law to software reigned until the early 1990s, after these cases, it became clear that copyright protection as applied to software would be quite limited.").

94. *See* 17 U.S.C. § 103 ("The subject matter of copyright as specified by section 102 includes compilations and derivative works, but protection for a work employing preexisting material in which copyright subsists does not extend to any part of the work in which such material has been used unlawfully.").

95. *See Feist*, 499 U.S. at 348.

96. David Sorkin, *Legal Problems in Data Management: IT & Privacy at the Forefront: "Big Data": Ownership, Copyright, and Protection*, 31 J. MARSHALL J. INFO. TECH. & PRIVACY L. 565 (2015) (stating that this kind of protection "would be important if someone took your entire database. If you collected— if someone copied U.S. News and World Reports' rankings of all the law schools with all their data in the columns and the way that they are

arrangement process were automated. The Compendium of U.S. Copyright Office Practices specifies that "the Copyright Act protects 'original works of *authorship*" and that to qualify as a work of authorship a work "must be created by a human being."[97] Thus, if the arrangement of the data is automated, then copyright protection likely won't attach at all.[98] This same dilemma arises in the context of copyrighting software. While copyright protection does extend to computer programs, only those portions of the software authored by humans will be protected by copyright.[99] This provides very narrow protection for AI-powered software.

Because protection under patent and copyright law is so limited in scope, where available at all, technology companies use trade secrecy to protect both datasets used to create the algorithmic models used in AI-powered software and the algorithmic models themselves.[100] In the United States, trade secret protection flows from both state and federal law with each state adopting its own trade secrets regime.[101] While the definition of trade secret may not be completely uniform under each of these regimes, certain elements are

---

arranged and just republished it as their own, that would be copyright infringement, for sure. Not just because of the images or the formatting of the page, but because the compilation is protectable by the way that they selected and arranged that data.").

97. U.S. Copyright Office, *Copyrightable Authorship: What Can be Registered, Compendium of U.S. Copyright Office Practices*, 21 (2021); *see Feist*, 499 U.S. at 346 (specifying that copyright law only protects the fruits of intellectual labor that are founded in the creative powers of the mind); Robert C. Denicola, *Ex Machina: Copyright Protection for Computer-Generated Works*, 69 RUTGERS UNIV. L. REV. 251, 264-70 (2016) (providing a more extended discussion of the copyrightability of computer-generated works).

98. Note, there has been a movement suggesting that requiring human being creators should be eliminated and some progress has been made in that regard in the patent sphere abroad. *See Recent Developments in Artificial Intelligence and IP Law: South Africa Grants World's First Patent for AI-Created Invention*, NATL. L. REV., (Aug. 3, 2021) https://www.natlawreview.com/article/recent-developments-artificial-intelligence-and-ip-law-south-africa-grants-world-s (discussing the granting of the world's first patent on an invention created by an AI inventor).

99. NIMMER ON COPYRIGHT § 2A.10 1 ("The Computer Software Copyright Act of 1980 adds to the Copyright Act an explicit definition for 'computer program.' Its legislative history specifies that the amendment 'has the effect of clearly applying the 1976 law to computer programs … .' That amendment dispels any lingering doubts as to the copyrightability of computer programs. It is therefore now firmly established that computer programs qualify as work of authorship in the form of literary works, subject to full copyright protection.").

100. *See* Steve Lohr, *Google Schools its Algorithm,* N.Y. TIMES, Mar. 5, 2011, at WK4 https://www.nytimes.com/2011/03/06/weekinreview/06lohr.html (discussing Google's algorithm and noting that Google elects to protect its algorithm as a trade secret).

101. Michael J. Kasdan, Kevin M. Smith, & Benjamin Daniels, *Trade Secrets: What You Need to Know*, THE NATL. L. REV. (Dec. 12, 2019) https://www.natlawreview.com/article/trade-secrets-what-you-need-to-know ("There is no uniform definition of "trade secret," because trade secrecy law developed at both the state and federal level. Historically, protection of trade secrets was a matter of state law, with each state developing its own definitions and rules.").

shared.[102] Whether datasets qualify for trade secret protection likely hinges on the publicness of the dataset and the measures taken by the company to keep the dataset secret.[103]

Datasets used by tech companies to create AI-powered software can be public or private. For example, if a company wanted to create an algorithmic model that predicts which kinds of people are likely to purchase a home in a particular area, they could rely upon public census data such as the American Community Survey Demographic and Housing Estimates.[104] In contrast, if a company running a search engine wanted to create an algorithmic model that could predict harmful interactions between pharmaceuticals, they could rely on consumer search data that is not publicly available.[105] Private datasets, so long as they remain secret, are strong candidates for trade secret protection.[106] If the initial dataset is public, however, then trade secret protection likely only attaches to the training data resulting from the data mining process.

Recall, however, that secrecy is a requirement for trade secret protection, and courts will evaluate whether tech companies took reasonable measures to maintain the secrecy of their datasets.[107] Whether certain courses of action qualify as reasonable depends largely on the nature of the information and the industry.[108] Generally, satisfying this standard would require that companies guard against known risks such as employees downloading the datasets onto portable devices,[109] and limit employee knowledge and access to only those that absolutely need to know.[110]

---

102. *Id.* ("While the number of definitions continue to multiply as federal courts have gotten involved, every definition shares a few key elements.").

103. *See* Amit Jaju. *How to Protect your Trade Secrets and Confidential Data*, ECON. TIMES, Mar. 5, 2022, https://economictimes.indiatimes.com/small-biz/security-tech/technology/how-to-protect-your-trade-secrets-and-confidential-data/articleshow/90010269.cms?from=mdr (discussing steps businesses can take to protect data as a trade secret).

104. *ACS Demographics and Housing Estimates, American Community Survey*, U.S. CENSUS BUREAU, https://data.census.gov/table?q=19701&tid=ACSDP5Y2021.DP05.

105. *See, e.g.,* Michael Mattioli, *Disclosing Big Data,* 99 MINN. L. REV. 535, 540 (2014) (describing the 2010 Stanford drug study wherein researchers at Stanford and Columbia partnered with Microsoft to develop a new way to predict harmful interactions between pharmaceuticals by analyzing millions of online searches made on several search engines which become the initial dataset for their study).

106. *Id.* at 550 (arguing that information-based processes that are not readily perceived by consumers are particularly well suited for trade secret protection).

107. *Id.* (stating that information need only be the subject of reasonable efforts to prevent disclosure to merit trade secret protection).

108. *See* Dupont Denemours & Co. v. Christopher, 431 F.2d 1012, 1016 (5th Cir. 1970) (explaining that reasonable measures does "not require the discoverer of a trade secret to guard against the unanticipated, the undetectable, or the unpreventable" to qualify).

109. *See* Waymo LLC v. Uber Techs., Inc., 870 F.3d 1350, 1355 (Fed. Cir. 2017) (describing an employee that downloaded thousands of documents related to Waymo's driverless vehicle technology).

110. VAN LINDBERG, INTELLECTUAL PROPERTY AND OPEN SOURCE, 119-131 (2008) (describing Coca Cola's famous technique of limiting of the formula to roughly two people and keeping the only original copy of the formula in SunTrust bank's main vault in Atlanta).

Protecting the algorithmic model generated from these datasets follows the same requirements. Courts will evaluate whether tech companies took reasonable measures to maintain the secrecy of their algorithms. The process by which computers take inputs, the training data, and transform it into output data, the algorithmic model, is commonplace.[111] Anyone with access to the training data and a basic understanding of the machine learning process would be able to generate the resultant algorithmic model. Trade secrecy does not protect against independent development.[112] Thus, tech companies looking to protect their algorithmic models as a trade secret would be wise to hold the dataset as a trade secret as well to guard against independent development. Of the three IP regimes discussed here, trade secrets offer the best protection for tech companies looking to exploit, commercially or otherwise, datasets and the algorithmic models used in AI-powered software.

Ultimately, this means that our current IP regime as applied to data and algorithmic models provides a strong incentive for tech companies whose business models rest on data and algorithms to avoid transparency. To maintain trade secret protection, technology companies must take any and all reasonable efforts to keep such information secret. Further, once such information has been shared with the public, trade secret protection is lost.[113] Any transparency as to the initial data used, training data used, class labels, or target variables risks devaluing the trade secret or complete loss altogether.[114]

This lack of transparency is significant because it presents a barrier to third-party auditing. Presently, our only indication of whether an algorithmic model exhibits bias is to look at its results. Joy Buolamwini, for example, was working with facial analysis software when she noticed that the software was unable to identify a broad range of skin tones and facial structures.[115] Companies are loath to hand over initial data, training data, class labels, target variables, or even the details of the algorithmic models or resultant software, which is exactly what third parties would need to examine in order to audit

111. *See* Jory Denny, *What Is an Algorithm? How Computers Know What to Do With Data*, THE CONVERSATION (Oct. 16, 2020), https://theconversation.com/what-is-an-algorithm-how-computers-know-what-to-do-with-data-146665 (describing in plain terms the machine learning process and noting that it is commonplace for things like recommendations, predictions, and looking up information).

112. *See Frequently Asked Questions: Trade Secrets*, WIPO (2022), https://www.wipo.int/tradesecrets/en/tradesecrets_faqs.html (noting that a trade secret owner cannot stop others from using the same technical or commercial information if they acquired or developed such information independently by themselves through their own research and development, reverse engineering, or marketing analysis, etc.).

113. *See* Tom Kulik, *NDAs & How to Lose Your Trade Secrets Without Really Trying*, ABOVE THE LAW, (Dec. 11, 2018), https://abovethelaw.com/2018/12/ndas-how-to-lose-your-trade-secrets-without-really-trying (stating that disclosure can cause a loss of status as a trade secret).

114. *Id.*

115. Janine Liberty, *How I'm fighting bias in algorithms*, MIT MEDIA LAB (Mar. 29, 2017), https://www.media.mit.edu/posts/how-i-m-fighting-bias-in-algorithms.

an AI system for potential biases.[116] Creating a fifth category of IP, however, one that applies specifically to AI and data and employs a human rights framework could disrupt this dynamic.

At the ideation stage, companies preparing to take on investment would be counseled to bring their AI system into compliance with this human rights framework in an effort to entice investors, much like companies already do with patent rights as discussed above. At the investment stage, investors would be heavily incentivized to pour money into companies whose technology complied with this human rights framework so long as it provided stronger protection than trade secret protection. And ultimately, consumers would benefit from AI that was built with equity in mind.

B. *Creating a Diverse Investment Environment*

The demographic makeup of both VC firms and founding teams matter for directing the flow of investment dollars to equitable AI. There are many investment funds in the United States actively investing in AI in young companies and nascent industries. In the VC space, typically categorized by relatively smaller investments in younger companies or newer and unproven technologies, the past five years have seen tremendous growth. [117] According to PitchBook, over 4,600 investors have made at least one investment in the AI and ML vertical[118]since January 2016.[119] In 2021, the global total corporate investment in AI has reached almost $94 billion, a significant increase from the previous year.[120] It's clear that investment dollars are already flowing into this space. If we are to direct the flow of investment dollars specifically toward AI, one key driver is people.

---

116. *See* Jessica M. Meyers, *Artificial Intelligence and Trade Secrets*, A.B.A. (2019), https://www.americanbar.org/groups/intellectual_property_law/publications/landslide/2018-19/january-february/artificial-intelligence-trade-secrets-webinar (discussing business's interest in protecting proprietary information and how that impacts the ability of third parties to audit AI algorithms for biases).

117. Sindhu Sundar & Matthew Lynly, *19 Top Venture Capitalists to Know That Invest in AI and Machine-learning Startups Like Hugging Face and Databricks*, BUSINESS INSIDER, (Feb. 6, 2023), https://www.businessinsider.com/top-venture-capitalists-investing-ai-machine-learning-startups-2022-10. (stating that in 2021 venture capital deals in AI and machine learning amounted to $118 billion, a roughly 80% increase from 2020).

118. *See What are Industry Verticals?*, PITCHBOOK (2022), https://pitchbook.com/what-are-industry-verticals#:~:text=An%20industry%20vertical%2C%20however%2C%20is,from%203D%20printing%20to%20eSports (last accessed Sep. 28, 2022) (explaining a "vertical" is a group of companies that focus on a shared nice or specialized market spanning multiple industries).

119. *Id*.

120. Berger Thormundsson, *Global Total Corporate Artificial Intelligence (AI) Investment From 2015 to 2021*, STATISTA (May 19, 2022), https://www.statista.com/statistics/941137/ai-investment-and-funding-worldwide/#:~:text=In%202021%2C%20the%20global%20total,increase%20from%20the%20previous%20year.

Deal selection and deal sourcing are core to the VC business model. Research has shown that, when making decisions about which companies to back, VC firms place the greatest importance on the management or founding team.[121] Additionally, a VC firm's ability to generate or source a pipeline of high-quality investment opportunities is considered an important determinant of success.[122] This insight is key for understanding how demographics impact the flow of investment dollars because both these drivers for success focus on people. In terms of deal selection, VCs look for certain qualities in management teams – such as ability, industry experience, entrepreneurial experience, passion, and teamwork – and strategically focus on cultivating and selecting entrepreneurs particularly ones they had worked with in previous investments.[123] In other words, VC firms bet their business on being able to recognize successful teams. In terms of deal sourcing, the majority of VC deal flow comes from VC firms' networks both at the firm level and at the individual level.[124] Roughly thirty percent of a VC firm's deal flow is generated through the professional networks of individuals working for the VC, twenty percent are referred by other investors, eight percent are referred from a portfolio company, and thirty percent a proactively self-generated.[125] Most revealingly, research shows that very few founders who come to VCs seeking investment without any connection successfully secure such investment.[126] Essentially, VC firms operate as if the most valuable thing they do is connect with and recognize the right people.

This leaves the door wide open for bias. Research has indicated that VC firms tend to favor management teams that are similar to themselves in terms of type of training and professional experience.[127] Additionally, there is evidence that VCs tend to evaluate entrepreneurs who demonstrate similar decision-making processes, in other words who "think" like them, more favorably.[128] Research has also shown that our social networks tend to be dominated by people of the same race or ethnic background, with white Americans being the most likely to have more homogenous social networks.[129] Gender matters

121. Paul A. Gompers, Will Gornall, Steven N. Kaplan & Ilya A. Stebulaev, *How do Venture Capitalists Make Decisions?*, 135 J. FIN. ECON. 169, 170 (2020).
122. *Id.* at 175.
123. *Id.* at 178.
124. *Id.* at 175.
125. *Id.*
126. *Id.*
127. Nikolaus Franke, Marc Gruber, Dietmar Harhoff & Joachim Henkel, *What You Are is What You Like – Similarity Biases in Venture Capitalists' Evaluations of Startup Teams*, 21 J. BUS. VENTURING 802, 805 (2006).
128. Charles Y. Murnieks, J. Michael Haynie, Robert E. Wiltbank, Troy Harting, *'I Like How You Think': Similarity as an interaction Bias in the Investor-Entrepreneur Dyad*, 48 J. OF MGMT. STUD. 1533 (2011).
129. *See* Daniel Cox, Juhem Navarro-Rivera & Robert P. Jones, *Race, Religion, and Political Affiliation of Americans' Core Social Networks*, PRRI (Aug. 03, 2016), https://www.prri.org/research/poll-race-religion-politics-americans-social-networks.

in this context as well, as studies have shown gender differences can lead to corresponding differences in the extend of involvement in managerial networking.[130] If VC firms are demographically out of sync with the general population, then their deal selection and deal sourcing processes are likely to suffer from the biases above.

C. *A Deeper Dive into Demographics of VC Funds and Investments*

1. VC Fund Demographics

Currently, the VC firms that invest in AI companies and products do not reflect the demographics of the broader US population.[131] The investor population is largely overrepresented by male and White individuals. We pulled data on the top five hundred firms based on activity according to Crunchbase.[132] Crunchbase does not claim to have comprehensive information for all data fields, but actively provides updates and maintenance of the data.[133] Their data structures are also flexible and allow for multiple options for demographics, for example, they support over thirty gender types,[134] even if many categories do not have any records.

The race/ethnicity of individuals is not directly reported in Crunchbase, nor in any trustworthy data source we found, so we turned to our own predictive model to identify ethnic likelihood. Our process followed the methodology proposed by Imai, Kosuke, and Kabir Khanna.[135] Their method aims to reduce bias in individual-level ethnicity predictions by combining self-identified ethnicities from voter registration records with the Census Bureau's Surname List. Their method may have a lower bias rate than other ecological

130. *See* Eunju Rho & Kangbok Lee, *Gendered Networking: Gender, Environment, and Managerial Networking*, 78 PUB. ADMIN. REV. 409, 415 (2018).

131. *See QuickFacts*, U.S. CENSUS BUREAU https://www.census.gov/quickfacts/fact/table/US/PST045221 (last visited Sept. 28, 2022) (US Census Bureau QuickFacts provides statistics for all states and counties, and for cities and towns with a population of 5,000 or more.).

132. *See generally Crunchbase Rank (CB Rank)*, CRUNCHBASE, https://support.crunchbase.com/hc/en-us/articles/115010477187-Crunchbase-Rank-CB-Rank- (last visited Sept. 28, 2022) (Crunchbase rank uses proprietary algorithms to score entities and create a relative ranking of community visibility).

133. *See generally What is Crunchbase Rank and Trend Score*, CRUNCHBASE, https://about.crunchbase.com/blog/crunchbase-rank-trend-score (last visited Sept. 28, 2022) (Crunchbase indicates ranking may change weekly based on activities like fundraising, news, etc.).

134. *See generally Crunchbase Enterprise API*, SWAGGERHUB, https://app.swaggerhub.com/apis-docs/Crunchbase/crunchbase-enterprise_api/1.0.3#/Person (expansion of the "gender" model box displays enumerated values possible through the API) (last visited Sept. 28, 2022) (Crunchbase data schema provides possible values for the database items.).

135. *See* Kosuke Imai & Kabir Khanna, *Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Record*, 24 POL. ANALYSIS 263, 264–66 (2016).

methods, especially for minorities, and claims to have a higher true positive rate.[136]

TABLE 1: INVESTOR DEMOGRAPHICS BY GENDER

	% Male	% Female	% White	% Asian	% Hispanic	% Black
Investors in AI	79.8%	19.8%	80.0%	16.0%	1.7%	1.3%
All Investors	81.5%	18.0%	81.4%	14.8%	1.8%	1.2%
US Population[137]	49.2%	50.8%	76.3%	5.9%	18.5%	13.4%

Note: Percentage of gender may not equal 100% as some investor genders are not specified.

By far the biggest divergence in demographics by investors is by gender, followed by the underrepresentation of Hispanic, Black, and finally the overrepresentation of Asian investors. If we compare race/ethnicity by gender, we see a similar breakdown as both genders carry the same under- and over-representations by race/ethnicity. The distributions are broken out by investor type, AI vs. All, and look similar. We think this speaks more to the proliferation of AI investments.

TABLE 2: INVESTOR DEMOGRAPHICS BY RACE/ETHNICITY AND GENDER

		% White	% Asian	% Hispanic	% Black
All Firms	Male Investors	81.9%	14.6%	2.2%	1.3%
AI Investors	Female Investors	75.2%	19.8%	3.0%	2.0%
All Firms	Female Investors	75.7%	19.3%	3.1%	1.9%
AI Investors	Male Investors	81.5%	15.0%	2.1%	1.3%

In the venture community, certain funds and investors are viewed as thought leaders and investment trend setters[138] This could be establishing new thematic investment categories, guiding founders (more details in the subsequent sections), structuring financing, or even the makeup of the firm itself. We looked at the demographics of the top 10% of firms,[139] representing the

136. *Id.* (The methodology from Kosuke Imai and Kabir Khanna may have a lower bias rate than alternative methods because individuals self-report both their ethnicity and name together. This may create a more accurate picture of the true ethnicity distribution for names, for example when a single surname can cross ethnic backgrounds (e.g., "Johnson").).

137. U.S. CENSUS BUREAU, *supra* note 131 (providing statistics for all states and counties, and for cities and towns with a population of 5,000 or more).

138. *See generally The Midas List: The World's Best Venture Capital Investors in 2022*, FORBES, https://www.forbes.com/midas/ (last visited Sept. 28, 2022) (A small example of trendsetters Forbes Midas List, which annually ranks global venture investors. In 2022 their list of top 100 global investors resided at 63 investment firms.).

139. CRUNCHBASE, *supra* note 132 (Crunchbase rank uses proprietary algorithms to score entities and create a relative ranking of community visibility.).

top fifty, in our sample and see similar demographic breakdowns. Though the over/under-representation is somewhat muted, the breakdowns reflect a similar pattern. Interestingly, forty-six of the fifty had at least one investment in an AI company in the past five years.

TABLE 3: INVESTMENT PATTERNS OF THE TOP FIFTY VENTURE CAPITAL FIRMS BY GENDER

		Total	% White	% Asian	% Hispanic	% Black
Top 50 Firms	Male Investors	72.1%	79.3%	16.8%	2.4%	1.5%
Top 50 Firms	Female Investors	27.9%	71.7%	23.8%	2.9%	1.7%

2. VC Fund Demographics

Our analysis continued to explore the makeup of company founders to identify and measure demographic skews in leadership. Crunchbase curates a database of company leadership often including the founder or founding team and other early and senior employees.[140] While we don't believe the dataset is a complete set, the average number of employees listed per company is 2.5 and indicates founders are likely listed.[141] We followed the same process to identify the race/ethnicity of founders as we did of investors. Founder gender was not specified, and we are unable to predict a binary gender through names.

TABLE 4: ETHNICITY OF VENTURE-FUNDED COMPANY LEADERSHIP

	% White	% Asian	% Hispanic	% Black
All Companies	66.4%	16.3%	2.9%	1.3%
AI Companies	70.5%	20.0%	3.0%	1.2%

Note: *"All companies" is the aggregation of all companies receiving venture funding since 2016.*

140. *See generally What are the Guidelines for Adding Content to Crunchbase?*, CRUNCHBASE, https://support.crunchbase.com/hc/en-us/articles/115011260487-What-are-the-guidelines-for-adding-content-to-Crunchbase- (last visited Sept. 28, 2022) (Crunchbase allows people to edit the database, but it is actively moderated by the Crunchbase team.).

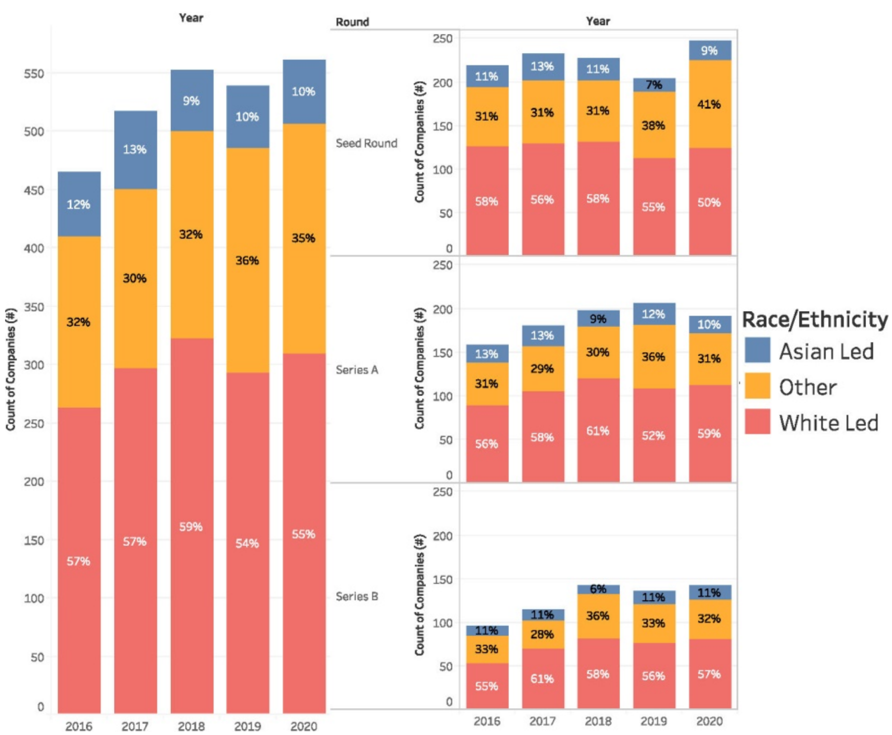
141. Based on Crunchbase data pulls querying company profiles and all people with a known relationship to the company (i.e., often, but not always, founders or employees) which is also stored in the Crunchbase database (i.e., relationship information from other sources like LinkedIn is not included). Queries est. 07/2021 – 08/2021.

3. Early Stage AI Companies

To get a better estimate of the amount of funding in new AI systems, we focused on early-stage funding rounds, specifically Seed, Series A, and Series B, of AI-only companies since 2016. We think the demographic trends of investment are important.

Please note, in the tables and images below the race/ethnicity categories have changed to be White-led, Asian-led, and Other. The Other category includes companies led by other individual race/ethnicity groups like Black and Hispanic and includes mixed leadership like those that are White- and Asian-led. The Other category was created to help minimize sparse tables and images, where many categories had values of zero in the different groupings.

FIGURE 1: FUNDING COUNTS BY RACE/ETHNICITY BY YEAR

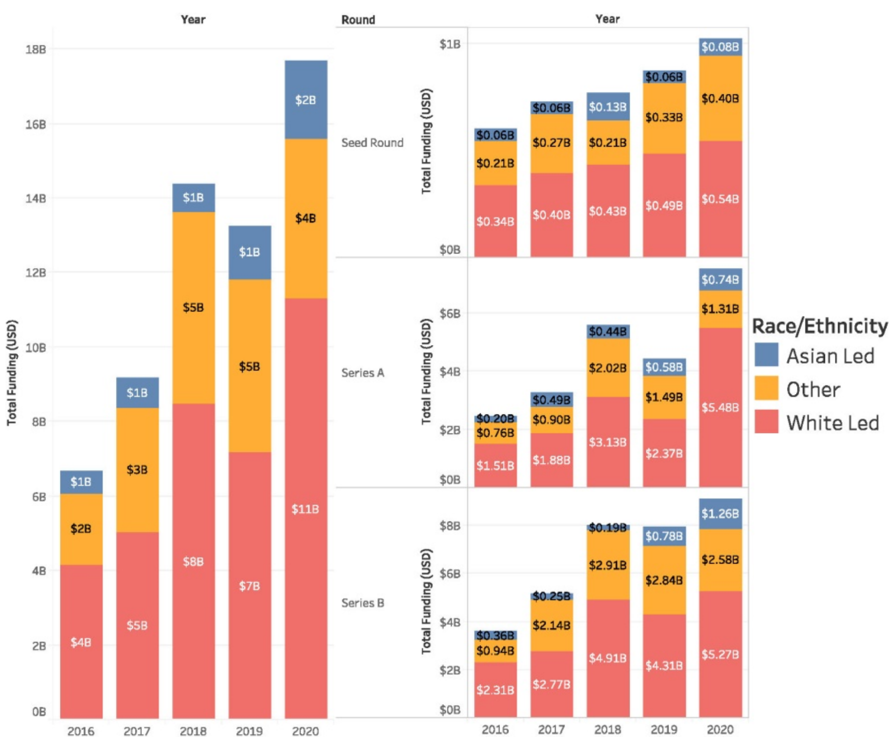


<u>Note:</u> A White or Asian-led company means the company had exclusively White or Asian founders. Other leadership includes Hispanic, Black, or mixed leadership. Companies are exclusively those categorized as operating AI & Machine Learning or Data and Analytics. Funding rounds are Seed, Series A, and Series B.

As seen above (Figure 1), since 2016 most AI companies receiving venture funding by both dollars and count had all White founders. The second-largest group of companies had all Asian founders; however, this group received less funding than all other groups, including Black, Hispanic, and mixed founders, where mixed founders may also consist of White and Asian only.

Since 2016 the total number of AI companies receiving funding has largely increased with roughly 20% more companies receiving early funding in 2020 compared to 2016. (See Figure 1). However, the percentage of funding by ethnicity has remained constant within a margin of +/- 2% over the years. Companies with all White founders make up more than 55% of companies funded each year since 2016. If we compare these numbers back to general population, the skew changes slightly to overrepresenting Asians. Compared to the overall population, White founders are under-represented.

FIGURE 2: FUNDING AMOUNTS BY RACE/ETHNICITY BY YEAR



<u>Note:</u> Funding is the aggregate sum of dollars received by companies from Venture funds in early funding rounds (Seed, Series A, Series B).

In terms of total dollars, White-led companies received the bulk of the total investment and have received most of the funding since 2016. In every year since 2016, White-led teams have received at least 54% of total funding. White-led teams also had the highest funding percentage in each early round of funding. (See Figure 2).

However, it is also important to analyze the funding received by each team, since each acts independently. If we look at the median funding per AI company broken out by funding round, see table below, we see a slightly different story, where white-led teams are not always the highest funded. This paints a slightly different picture showing that diverse teams are funded at similar levels to their non-diverse counterparts. Taken together with the aggregate funding, the number of diverse teams is still underrepresented.

TABLE 5: FUNDING AMOUNTS BY ROUND AND RACE/ETHNICITY

	Race/Ethnicity	**2016**	**2017**	**2018**	**2019**	**2020**
Seed Round	**Asian Led**	$3M	$3M	$5M	$5M	$3M
	Other	$2M	$3M	$4M	$4M	$4M
	White Led	$3M	$2M	$3M	$4M	$4M
Series A	**Asian Led**	$9M	$15M	$15M	$20M	$16M
	Other	$11M	$12M	$18M	$14M	$15M
	White Led	$12M	$12M	$12M	$14M	$18M
Series B	**Asian Led**	$30M	$17M	$22M	$29M	$65M
	Other	$19M	$34M	$34M	$31M	$42M
	White Led	$22M	$30M	$38M	$30M	$39M

Note: The increase in median funding both by round and by year is expected. Table cells in yellow highlight the race/ethnic group with the highest median funding per category. For example, in Series B fundings in 2018, "White Led" companies received the largest median funding.

4. Additional Analysis of Funding

Tracking the Seed round may offer a closer look at early ventures in new AI concepts, as opposed to later rounds of funding which require a more proven track record. Seed funding in the last few years shows a good mix of median funding by the demographic of the team. It also shows a net increase in the total amount of teams that were funded from diverse backgrounds.

Another analysis that could be useful is looking at "outsized funding" rounds. An outsized funding round is one where a company receives more funding than is common for a company at their stage of development.[142] We define outsized as funding with a Z-Score >= 3, or greater than 99% of all category peers, and view this as taking a larger financial risk on the company. Although there is no specific formula for investment amounts, it might be revealing to see if a specific demographic had outsized funding. Investment sizes may depend on many factors including the industry, market, traction so far, funding team, and investor ability to fund. Although White-led teams had the highest count and percentage of outsized funding, their percentage was slightly lower than the total amount of funding the group received.

TABLE 6: NUMBER OF TEAMS RECEIVING OUTSIZED FUNDING BY RACE/ETHNICITY

Race/Ethnicity	Count of Outsized Funding	Percentage of Outsized Funding
Asian Led	5	12%
Other	15	36%
White Led	22	52%

TABLE 7: NUMBER OF TEAMS RECEIVING OUTSIZED FUNDING BY RACE/ETHNICITY AND YEAR

	Race/Ethnicity	**2016**	**2017**	**2018**	**2019**	**2020**
Seed Round	**Asian Led**	[redacted]	[redacted]	1	[redacted]	[redacted]
	Other	1	4	[redacted]	4	0
	White Led	2	3	2	3	2
Series A	**Asian Led**	[redacted]	[redacted]	[redacted]	[redacted]	2
	Other	[redacted]	[redacted]	1	[redacted]	1
	White Led	[redacted]	[redacted]	[redacted]	[redacted]	1
Series B	**Asian Led**	[redacted]	[redacted]	[redacted]	1	1
	Other	[redacted]	2	1	1	[redacted]
	White Led	1	[redacted]	2	3	3

142. A company's stage of development can be defined and measured in numerous ways such as age, product maturity, revenue growth, etc. Many companies seek investment funding at specific benchmarks like $1M or $10M in total annual revenue and investors use these benchmarks as comparables. As such, we consider the stage of company development to correspond to the round of financing the company is seeking, Seed, Series A, or Series B.

Note: *Outsized funding is when a company has a funding round with a z-score >=3 based on the year and round. For example, there was only one company with an outsized Series B round in 2016 and it was White-led.*

That VC firms and founding teams in the AI space are out of sync with the demographics of the broader U.S. population is relevant because diverse teams increase the likelihood that the AI they build will be equitable. Recall, for our purposes, that we are relying upon a human rights framework to operationalize our conception of equity in this space. We posited that if AI is to be equitable it must, at a minimum, avoid violating internationally recognized human rights. Biases like those discussed in Part I of this article can be powerful drivers for creating AI that is violative of international human rights. For example, article eleven of the United Nations Declaration of Human Rights states that "[e]veryone charged with a penal offence has the right to be presumed innocent until proved guilty according to law in a public trial at which he has had all the guarantees necessary for his defense." AI-powered predictive recidivism risk assessment tools such as COMPAS purport to identify whether a given person is likely to re-offend, and these tools have been used to determine whether an individual should be detained pretrial or released on parole.[143] Troublingly, evidence has emerged that these risk assessment tools exhibit anti-Black biases similar to those discussed in Part I. Research has shown that these tools systematically miscalculate the risk of recidivism for individuals from minority communities, wrongly identifying minority individuals as likely to reoffend at a rate much higher than those from majority communities.[144] The results pose a risk of violating the human rights of these defendants. Because these tools' predictions are not subject to meaningful review – due in part to judicial capacity and the objective veneer associated with the AI's outputs – miscalculations adversely impact the defendants' right to a fair public trail, to a defense, and to an appeal.[145] Research has indicated that there is a "substantial risk that the rights of minority groups to equality and non-discrimination will be adversely affected by such tools."[146]

Industry and thought leaders alike have suggested that the key to mitigating bias in AI is to hire more diverse teams, or short of that to bring diversity

143. *See* Raso et al., *supra* note 17, at 21; Mirilla Zhu, *An Algorithmic Jury: Using Artificial Intelligence to Predict Recidivism Rates*, 92 YALE SCI., Sept. 2020, at 22 (identifying COMPAS as one such risk assessment tool and briefly describing the adoption and use of risk assessment tools.).

144. *See* Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (finding that a recidivism risk assessment algorithm was twice as likely to incorrectly label black defendants as future criminals as compared to white defendants.); *See also* Raso et al., *supra* note 17, at 23.

145. *See* Raso et al., *supra* note 17, at 22.

146. *Id.* at 23.

into the process in other ways.[147] Experts have argued that more diversity within the AI community would entail an increased ability to anticipate, review, and spot bias and engage the communities affected.[148] Thus, ensuring that investment dollars flow to diverse teams is one way to increase the likelihood that the AI developed will be equitable. As we've established, the investment dollars currently flow towards teams that do not reflect the demographics of the broader U.S. population. Given the ways in which VC firms source and select deals, this is unsurprising. If we are to redirect the flow of investment dollars towards more diverse teams, ones that do reflect the demographics of the broader U.S. population, then we must also encourage diversity among VC firms as well.

We have argued for two ways to guide the flow of investment dollars towards equitable AI. The first is a direct approach: use a new category of IP to condition ownership of data and AI upon compliance with a human rights framework. The second is an indirect approach: create a diverse investment environment, thus increasing the likelihood that the AI developed by founding teams will be equitable. Doing either in isolation likely is not robust enough to have durable impact. By supplementing the direct approach with our indirect approach, we are creating an ecosystem that has multiple levers for incentivizing equitable AI rather than just one.

D. Analyzing Public Statements Made by VC Funds Addressing Diversity and Inclusion

As mentioned above, many VC firms made Diversity and Inclusion (D&I) commitments in the summer of 2020 during the resurgence of the Black Lives Matter movement and the heightened awareness of racial inequality in the United States.[149] That summer, firms like Sequoia Capital, Bain Capital, and The Carlyle Group pledged support for racial justice, promising to dedicate themselves to efforts to close the representation gap in their firms

---

147. *See* Kathy Baxter, *What Is AI Bias Mitigation, and How Can It Improve AI Fairness?*, INFOWORLD (Aug. 24, 2021), https://www.infoworld.com/article/3630450/what-is-ai-bias-mitigation-and-how-can-it-improve-ai-fairness.html ("The likelihood of negative outcomes increases with a lack of diversity in the teams responsible for building and implementing the technology."); Veronika Shestakova, *Best Practices to Mitigate Bias and Discrimination in Artificial Intelligence*, 60 PERFORMANCE IMPROVEMENT 6, 9 (July 6, 2021) (identifying lack of diversity in teams as one of the challenges for mitigating bias).

148. *See* James Manyika, Jake Silberg & Brittany Presten, *What Do We Do About the Biases in AI?*, HARV. BUS. REV. (Oct. 25, 2019), https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai.

149. *See* Arielle Pardes, *VC Pledged to 'Do Better' on Diversity. It's Barely Changed*, WIRED (June 10, 2021), https://www.wired.com/story/vc-pledged-better-diversity-its-barely-changed (discussing the proliferation of diversity commitments following George Floyd's murder). *See generally* Larry Buchanan, Quoctrung Bui & Jugal K. Patel, *Black Lives Matter May Be the Largest Movement in U.S. History*, N.Y. TIMES (July 3, 2020), https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html (discussing the enormous scale of protests in the wake of George Floyd's murder, and their social impact).

and commit to transformational measures geared towards increasing D&I.[150] While many of these commitments were public, only a few firms have published statistics regarding their company diversity numbers, including investments and results, and data is not readily accessible.

We researched public statements made by fifty prominent VC and PE firms to understand different approaches to publicity, namely publicly supporting racial justice and promoting diversity. However, as there is no single repository of statements, we scoured all public comments and channels to identify if messages were related to racial equality. Our analysis found no standard pattern or practice in making public comments and found certain methods conveyed a stronger message and commitments.

We found that each firm took a different approach to its commitments on diversity, including the absence of public-facing statements. Our analysis determined that each firm's chosen method contributed as much to the message as the statement itself. In other words, we felt the medium each firm chose was a factor in measuring the overall strength of the statement. In addition to medium, other factors we assessed included i) characteristics of the statement, the statement type, and if the statement came from a specific person or team, ii) target groups that were specifically mentioned in the statement, if any, iii) any internal commitment that the firm would make, iv) an external commitment made by the firm, including any specific financial commitment, and v) any follow-up to stated commitments found after the original posting. Below, we explore the different factors we considered in assessing public statements by prominent funds.

We considered different aspects of the statement itself, including the channel it was published, the type of statement released, and the signature(s) listed. These three aspects reflect true ownership over the statement, though we found that any posting at all created a positive contribution. Many firms posted on their social media profiles (e.g., Twitter), which garnered many views and allowed user feedback, but these posts were eventually pushed out of sight by new posts or other information. Others achieved more longevity by posting directly to their website, on the landing page, a firm blog, or a

---

150. *See* Diann Lawson, *Sequoia is Committed to Diversity, Equity and Inclusion*, SEQUOIA (Dec. 3, 2020), https://www.sequoia.com/2020/12/our-commitment-to-diversity-equity-and-inclusion (asserting Sequoia's commitment to diversity and inclusion); *Then. Now. Next. 2022 Diversity, Equity, and Inclusion Report*, BAIN & COMPANY (2022), https://www.bain.com/about/further-global-responsibility/diversity-equity-inclusion/dei-report (noting that Bain made a commitment in June 2020 and assessing its progress); *Bain's Commitment to Promote Racial and Social Equity*, BAIN & CO., https://www.bain.com/about/further-global-responsibility/diversity-equity-inclusion/bains-commitment-to-promote-racial-and-social-equity (last visited Sept. 28, 2022) (asserting Bain's commitment to promoting racial and social equity); Anne Sraders, *Private Equity Wields a Lot of Power to Promote Diversity. Here's How Carlyle Is Upping Its Efforts*, FORTUNE (Aug. 3, 2020), https://fortune.com/2020/08/03/private-equity-wields-a-lot-of-power-to-promote-diversity-heres-how-carlyle-is-upping-its-efforts (discussing The Carlyle Group's commitment to promoting diversity).

dedicated firm space. Firms also differed in who released the statement or signed it, such as an executive, the D&I chair, or simply to be understood to be from the entire team. Although firms greatly vary in size, we felt a statement with individual names conveyed stronger ownership. A statement signed by the CEO or executive is a notable case because it portrayed that the firm's leadership was committed to D&I initiatives.

Within the text of the statement, some firms directed their actions to specifically named groups like Black Lives Matter, LGBTQ, Women, and other minority groups. Firms that did this often related the statements to "injustices" as a whole. Naming groups also allowed them to demonstrate the intersectionality between groups, existing behaviors, and what actions might be necessary to overcome them. We felt this was an important factor because it reveals an awareness of the situation and scope of the problem that groups encounter.

Many firms referenced internal-facing changes that involved development programs or other initiatives like diversity recruitment or increasing diversity in leadership. For example, some firms used the opportunity to announce a new D&I position, chair, or partner with specific duties at the firm. Other firms promised to improve diverse internal representation on their portfolio companies' boards and management teams. Venture firms can make a direct impact through financing, but they also have many other levers to pull, such as recommending executives and advisors to young companies. We felt both actions by firms showed a strong commitment as they both translate to increasing diverse perspectives in influential roles within companies.

Firms also made commitments to external-facing changes, recognizing that they can significantly impact their communities and the economy directly. We measured different factors of external commitments like action steps, community development, and specific financial promises. An example of all three came from Andreessen Horowitz with their Talent X Opportunity Fund, designed to help entrepreneurs who historically lack resources to succeed. Other funds made specific financial commitments to matching employee charitable giving or donating money to causes like legal defense funds.

Some funds took a fuller commitment to investing in diversity and created net new funds explicitly reserved for minority groups. Others, recognizing they did not have the resources to invest in minority groups, made funds of funds to support underrepresented general partners of funds[151] In other words, they found creative ways to support investments in underrepresented groups even if they didn't make direct investments in diverse companies. Companies that commit to donating a specific sum of money, similar to

151. *See, e.g.*, *About*, SCREENDOOR, https://www.screendoorpartners.com/about (last visited Sept. 28, 2022) (describing the mission of Screendoor, an investment fund dedicated to funding other diverse fund managers).

matching, demonstrate fidelity to social equality and justice by going beyond press releases to actionable monetary steps.

After analyzing all statements, we determined specific features of public statements portrayed a greater outward commitment to D&I. Since investment firms vary in size, focus, and process, we don't think there is a "one size fits all approach" and don't think it is reasonable to expect every firm to perform or comment the same way. We believe assertive statements show higher ownership, knowledge, and intersectionality of the groups and issues while also demonstrating more creative ways to overcome these challenges. When crafting a "toolkit" that standout firms have done, some helpful "instruments" share a focus on long-term outcomes through internal and external actions. For example, a long-term external action is providing financial support (e.g., employee matching programs, distributing $X sum of money over some time – such as Bain Capital providing more than $100 million from partners over the next ten years to nonprofit organizations focused on social justice, civil rights, and racial equality while also doubling their matching gifts to these causes done by employees), and committing to improving diverse representation with goalposts (e.g., a percentage or number requirement of diverse board members or executive officers – like Techstars's goal of having at least 1,000 diverse CEOs go through their accelerators by 2026). Compared to momentary relief or statements, initiatives that commit to structural adjustments have the most lasting and significant impact. Because a "one size fits all approach" is suboptimal, in assessing whether a firm has established a sufficient commitment statement we should use a totality of the circumstances standard evaluating along the factors discussed above.

We conducted the second portion of the analysis on statements nearly a year later, in the summer of 2021. The goal was to measure progress towards each commitment and ensure steps were taken after public statements. As of August 2021, there is some evidence to suggest that VC's have made progress following commitment statements including increased diversity of hires, continued statements supporting D&I, and increased D&I programming.[152] Others would argue, however, that progress relative to commitment is lacking and that these commitments ultimately lack the teeth needed to drive real change.[153] For greater traction, and in order to incentivize VC firms to diversify, there needs to be an initiative to standardize reporting and make publication compulsory. Firms should be motivated to promote diversity and set diversity goals within their investment portfolio. Additionally, providing a

152. *See* Rangita de Silva de Alwis, *New Research Finds Some Progress for DEI in Venture Capital and Private Equity*, THOMSON REUTERS (Sept. 13, 2021), https://www.thomsonreuters.com/en-us/posts/legal/diversity-pledges-venture-capital-private-equity (stating that progress has been made internally at the senior and junior investment professional levels within VC firms).

153. *See* Pardes, *supra* note 149.

blueprint for companies to improve their numbers is critical to support this goal.

III. OUR PROPOSAL: WHAT THE EQUITABLE ECOSYSTEM LOOKS LIKE IN PRACTICE

In Part I, we explored the issue of bias in AI systems. We provided a brief overview of how biases can leak into the AI development process and provided empirical evidence that such biases may already be manifesting in AI systems. In Part II we made the case for our two-pronged approach. We argued that a robust strategy called for both direct and indirect interventions, focusing on AI ownership structure for the former and the diversity of the investment system for the latter. In Part III, we endeavor to set out what this looks like in practice. Our goal in this section is to outline a) how our proposed fifth category of IP utilizing a human rights framework might look and b) how to create a diverse investment system.

A. *Exchanging Ownership for Compliance with Human Rights Framework*

We propose that this fifth category of IP rest on a central bargain: ownership in exchange for compliance with a human rights framework. Framing this category in terms of bargain would be consistent as each of the four existing categories of IP could be described in terms of bargain as well. For patents, the bargain could be ownership in exchange for disclosure.[154] For copyright, the bargain could be ownership in exchange for creativity.[155] For trademark, the bargain could be ownership in exchange for a reduction in consumer confusion due to counterfeiting or fraud.[156] And for trade secrets, the bargain could be ownership in exchange for increased innovation.[157] In each of these other categories, we the people grant ownership interests, or something akin to an ownership interest, in exchange for a societal good. This

154. *See* Shubha Ghosh, *Patents and the Regulatory State: Rethinking the Patent Bargain Metaphor After Eldred*, 19 BERKELEY TECH. L. J. 1315, 1319–21 (2004) (noting that the patent bargain concept is generally accepted in the patent law community).

155. *See* Raymond S. Ray, Jiayang Sun & Yiying Fan, *Does Copyright Law Promote Creativity? An Empirical Analysis of Copyrights Bounty*, 62 VAND. L. REV. 1669, 1671 (2009) (explaining that modern copyright law is based upon the theoretical tradeoff of protection in exchange for increased creative activity).

156. *See* Mark McKenna, *A Consumer Decision-Making Theory of Trademark Law*, 98 VA. L. REV. 67, 69–70 (2012) (arguing that courts understand trademark law's job to be to rid the marketplace of confusion); *What Is a Trademark?*, U.S. PAT. & TRADEMARK OFF., https://www.uspto.gov/trademarks/basics/what-trademark (last visited Sept. 28, 2022) (noting that trademarks help guard against counterfeiting and fraud).

157. *See* OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, *Statement of Administration Policy: S. 1890 – Defend Trade Secrets Act of 2016*, THE AMERICAN PRESIDENCY PROJECT (Apr. 4, 2016),

new category would extend that tradition, granting an ownership interest in exchange for developing and deploying AI in accordance with our proposed human rights framework.

To create our framework, we draw heavily upon the principles and themes outlined by the United Nations (UN) in its Guiding Principles on Business and Human Rights and the framework developed by Principles for Responsible Investment.[158] We draw upon these resources in creating our framework as they have already transfigured human rights principles such that they can be operationalized by the business sector. As a threshold matter, our framework endeavors to impose two primary obligations upon companies developing and deploying AI: i) to avoid causing or contributing to adverse human rights impacts through their AI technology, whether in development or deployment, and to address such impacts when they occur, and ii) to seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships[159] even if they have not contributed to those impacts.[160] This establishes both a negative and positive obligation. Note, we make no distinction between AI technology that causes or contributes to adverse human rights impacts through positive action or omission, and purposefully focus on impact as opposed to intention.[161] The animating principle underlying our new category of IP is social good, or human rights-respecting AI, in exchange for ownership. Allowing ownership to attach even if a company created AI systems or AI-powered services that are violative of human rights, either by omission or by mistake, particularly without providing for a remedy would frustrate this principle.

We propose that, for ownership to attach, businesses developing and deploying AI systems must conform to the following key tenets: i) policy commitment, ii) active inclusion, iii) due diligence, iv) access to redress, and v) transparency. Companies must craft a publicly available statement committing to upholding their responsibility to ensure their AI systems and AI-powered services respect human rights. Companies must also integrate active

158. U.N. Hum. Rts. Off. Of the High Comm'r, *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy Framework*, HR /PUB/11/04 (Jan. 2012), https://www.ohchr.org/documents/publications/guidingprinciples businesshr_en.pdf [hereinafter *UNGP*]; *Why and how investors should act on human rights*, PRINCIPLES FOR RESPONSIBLE INVESTMENT (Oct. 22, 2020) https://www.unpri.org/human-rights/why-and-how-investors-should-act-on-human-rights/6636.article (providing a helpful blueprint for how UNGP was applied to investor institutions).

159. *See generally* UNGP, *supra* note 159, at 15 ("'[B]usiness relationships' are understood to include relationships with business partners, entities in its value chain, and any other non-State or State entity directly linked to its business operations, products or services.").

160. *See id.* at 14–15 (discussing UNGP guiding principle 13, which closely tracks our proposed obligations).

161. *See id.* at 14 (advocating for corporate prevention or mitigation of harmful human rights impacts, regardless of the corporation's contribution).

inclusion[162] into their hiring and AI development processes. Companies must create due diligence processes to identify, prevent, mitigate, and account for how to address the impacts on human rights caused by its AI systems or AI-powered services. Companies must establish processes to enable remediation of adverse human rights impacts caused by their AI systems or AI-powered services or to which the company's AI or AI-powered services contribute. And lastly, Companies must commit to disclosing the techniques it uses to debias or otherwise ensure that their AI systems and AI-powered services respect human rights. We discuss each in more depth below.

The required policy statement is intended to serve as the nexus point for embedding into the company's AI development and deployment processes, from high-level strategy setting through its ultimate execution, a commitment to ensuring that the resultant AI system or AI-powered service conforms with the obligations outlined above. The policy statement should be approved at the most senior level of the company.[163] It should be informed by relevant internal and external expertise, as appropriate given the size of the company.[164] It should be publicly available and communicated internally and externally to all personnel, business partners, and other relevant parties.[165] It should be reflected in operational policies and procedures necessary to embed it throughout the AI development and deployment process.[166] It should stipulate the company's human rights expectations of parties directly impacted by its AI system or AI-powered service and should be revised from time to time in light of expanded use cases, applications, and business growth.[167] Additionally, companies should perform regular audits to measure compliance with their policy, the results of which should also be made public.

The active inclusion requirement is intended to ensure that companies are proactive in seeking a diversity of input when developing and deploying AI systems or AI-powered services. Companies should evaluate the populations likely to be affected by the output of their AI systems and AI-powered

---

162. *See* WORLD ECON. FORUM, HOW TO PREVENT DISCRIMINATORY OUTCOMES IN MACHINE LEARNING, GLOBAL FUTURE COUNCIL ON HUMAN RIGHTS 2016–18, 5 (Mar. 2018), https://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf [hereinafter WEF WHITE PAPER] (highlighting "active inclusion" as a core principle).

163. *See* UNGP, *supra* note 159, at 16 (stating that enterprises should express their commitment to embedding their responsibility to respect human rights into their business through a statement of policy that is approved at the most senior level of the business enterprise).

164. *See id.* (stating that policy commitments should be informed by relevant internal or external expertise).

165. *See id.* (stating that policy commitments should be made publicly available and communicated internally and externally to relevant parties).

166. *See id.* (stating that policy commitments should be reflected in operational policies and procedures necessary to embed it throughout the business enterprise).

167. *See id.* (stating that policy commitments should stipulate the enterprise's human rights expectations of personnel, business partners and other parties directly linked to its operations).

processes and endeavor to ensure those populations are reflected in the development process. Companies should actively strive to include folks from those affected populations in the development process, preferably through a combination of hiring and focus groups, and should consider the norms and values of those affected populations.[168]

The due diligence requirement is meant to ensure that companies are taking proactive steps to identify and mitigate negative outcomes as opposed to merely paying lip service to the matter of equity in AI. The due diligence standards of the Guiding Principles on Business and Human Rights provide an agreed set of norms and a shared language and institutional infrastructure to ensure that AI's potential for good can be met while holding the business ecosystem accountable to AI- related discrimination and bias.[169] Companies should take four core steps during the due diligence process. First, the company should identify actual and potential negative outcomes for people arising from the AI system or AI-powered service.[170] Companies should then prevent and mitigate the actual and potential negative outcomes identified.[171] Companies should track the ongoing management of human rights outcomes from their AI systems and AI-powered services.[172] And, companies should communicate to clients, beneficiaries, affected stakeholders, and the public about outcomes and the actions to be taken, if any.[173] The scope of a company's due diligence efforts should be in line with the size and complexity of the company's business, particularly the complexity of the AI system or AI-powered service, and should take into account the severity of potential human rights impacts.[174] Additionally, this due diligence should be ongoing and should evolve based upon changes in the human rights risks posed by the AI system or AI-powered service.[175] The risk assessment process should utilize a combination of internal and independent external human rights expertise as feasible depending on the size of the company and should involve meaningful

168. *See* WEF WHITE PAPER, *supra* note 163, at 12.

169. *See* UNGP, *supra* note 159, at 17 (establishing the parameters that businesses should follow in developing human rights due diligence efforts).

170. *See id.* (stating that the human rights due diligence process should include assessing actual and potential human rights impacts).

171. *See id.* at 16 (commenting that potential impacts should be addressed through prevention or mitigation while actual impacts that have already occurred should be addressed through remediation).

172. *See id.* at 17 (stating that the process should include tracking responses).

173. *See id.* (stating that the process should include communicating how impacts are addressed).

174. *See id.* at 18 (noting that human rights due diligence "[w]ill vary in complexity with the size of the business enterprise, the risk of severe human rights impacts, and the nature and context of its operations").

175. *See id.* (explaining that human rights due diligence should be "ongoing, recognizing that the human rights risks may change over time as the business enterprise's operations and operating context evolve").

consultation with the groups potentially affected by the AI system or AI-powered service.[176]

The access to redress requirement is intended to underscore companies' responsibility for the use and actions of their AI systems or AI-powered services.[177] We recognize that even with the best and most comprehensive processes and mitigation plans, negative outcomes and impacts are possible.[178] When a company has identified that it has caused or contributed to adverse impacts arising from its AI system or AI-powered services, the company should provide for or assist affected persons in accessing remediation through legitimate processes.[179] Requiring active engagement in remediation reaffirms the company's ultimate responsibility for outcomes regardless of intent or mistake.

The transparency requirement is intended to foster global collaboration in establishing effective techniques and protocols for creating human rights-respecting AI.[180] Due in part to our current IP system's incentivization of secrecy in this space, information about the specific methods used by companies to ensure their AI systems and AI-powered services are non-discriminatory and rights-respecting is not generally available.[181] Further, even if a company were to claim publicly that it utilized techniques to debias its AI systems and AI-powered services, third-party auditing of those techniques is likely not possible as most companies hold their data and AI as proprietary

176. *See id.* at 19 (suggesting that assessment of potential adverse human rights impacts should "draw on internal and/or independent external human rights expertise" and "involve meaningful consultation with potentially affected groups" as appropriate).

177. *See* WEF WHITE PAPER, *supra* note 163, at 5 (arguing that "leaders, designers and developers of ML systems are responsible for identifying the potential negative human rights impacts of their systems").

178. *See* UNGP, *supra* note 159, at 24 (acknowledging that even with the best policies and practices, a business enterprise may cause or contribute to an adverse human rights impact that it has not foreseen or been able to prevent).

179. *See id.* (stating that business enterprises should provide for or cooperate in remediation through legitimate processes when those businesses have identified that they have caused or contributed to adverse human rights impacts).

180. This step is inspired in part by the disclosure requirement imposed in patent law. Both scholars and justices have argued that disclosure is one of the benefits of the patent system because disclosure encourages second order innovation. *See* Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 481 (1974) (arguing that patent disclosures will stimulate ideas and the eventual development of further significant advances in the art and citing disclosure as a main goal of the patent system). *See also* Lisa Larrimore Ouellette, *Do Patents Disclose Useful Information?*, 25 HARV. J. OF L. & TECH. 546, 549 (2012) (arguing that bolstering the disclosure requirements for patents would increase their usefulness as a source of technical information for scientists that develop groundbreaking innovations).

181. *See* Ouellette, *supra* note 181, at 588 (arguing that businesses generally elect to protect data and AI as a trade secret and thus providing a strong incentive against transparency). *See also* Rowena Rodrigues, *Legal & Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities*, 4 J. OF RESPONSIBLE TECH. (2020) (noting the lack of algorithmic transparency and explaining its centrality in discussions about legal and human rights issues in the AI space).

information.[182] Compelling disclosure of techniques would ensure that others can learn from the successes and failures of AI development teams in their sectors.

B. *Creating a Diverse Investment Environment*

We are proposing a favorable tax adjustment on the gains made from successful investments championed by VC firms that are graded favorably on our commitment index. VC firms have already demonstrated a willingness to make commitment statements with respect to diversity, making commitments an attractive lever for incentivizing VC firms to increase diversity.[183] There is already industry buy-in for creating these commitment statements, our intervention is to create a system that turns this buy-in into tangible outcomes in terms of the demographic makeup of VC firms and founding teams by providing strong incentives for implementation. The commitment index coupled with tax adjustment is our proposed system.

The mechanics of our proposed system is briefly stated as follows: the favorable tax adjustment is the carrot is meant to motivate firms to strive for favorable grading on our commitment index,[184] and the commitment index grading process is designed to stimulate action on the part of firms to increase diversity among their ranks. Grading on the commitment index will be made public and will be in accordance with two key elements: i) does the VC firm have an optimal commitment statement with respect to diversity, and ii) has the VC taken robust steps to actualize that commitment. Optimal commitment statements should focus on long-term outcomes through internal and external actions and will be evaluated under a totality of the standard circumstances, and the audit process for determining progress should focus on concrete steps taken to advance that stated commitment. We determined what optimal commitment statements should focus on by researching recent efforts in the market.

Our proposed system is valuable because it is designed to incentivize parties at every juncture within the investment environment. The tax adjustment will allow VC firms that qualify for it to generate higher value for their

182. *See* Simson L. Garfinkel, *A Peak at Proprietary Algorithms*, AM. SCI., Nov.–Dec. 2017, at 326, 326–27, https://www.americanscientist.org/article/a-peek-at-proprietary-algorithms; Inioluwa Deborah Raji, Peggy Xu, Colleen Honigsberg & Daniel E. Ho, *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance*, 2022 AAAI/ACM CONF. ON AI, ETHICS, AND SOC'Y 557, 564 ("Lack of access to data and algorithmic systems strikes us as the most significant vulnerability of the current AI audit ecosystem.").

183. *See* Sraders, *supra* note 150 (VC firms have already demonstrated a willingness to make commitment statements with respect to diversity, making commitments an attractive lever for incentivizing VC firms to increase diversity.).

184. *See* Lourdes German & Joseph Parilla, *How Tax Incentives Can Power More Equitable, Inclusive Growth*, BROOKINGS (May 5, 2021), https://www.brookings.edu/blog/the-avenue/2021/05/05/how-tax-incentives-can-power-more-equitable-inclusive-growth (explaining how tax incentives can be a tool for influencing economic growth).

investors and themselves than VCs that do not, because the gains per deal will be taxed on a reduced basis. This value will complement the diversity dividend – the improved financial performance on measures such as profitable investments at the individual portfolio-company level and overall fund returns – that VC firms will gain from increasing diversity among their ranks.[185] It follows then, that diverse VC firms will be positioned to outperform non-diverse VC firms all else being equal. This is a powerful incentive because firms are interested in maximizing performance and they compete for investor capital. Diverse VC firms may be better positioned to win Limited Partners because they can argue that they deliver more value for investor dollars. We expect that, in light of VC firms' role in deal sourcing and deal selection as discussed above, this should have positive downstream effects on the diversity of the founding teams that secure investment. However, if the anticipated positive downstream effects on the diversity of founding teams do not seem to materialize, we could incorporate this aspect explicitly into two key elements. Essentially, we would require that VC firms include investing in diverse founding teams as part of their commitment statement and audit the concrete steps those VC firms take during the fiscal year to advance those commitments.

V. Conclusion

That AI will transform our world is undeniable; the decision facing lawmakers, technologists, and thought leaders now is whether we will ensure this transformation will be equitable for all. It is clear that theoretical discussions of bias in AI are being borne out in the market. There is already evidence that AI-powered recidivism predictors are potentially violative of certain minority communities' human rights. Additionally, evidence of anti-black bias in hiring algorithms suggests AI could be facilitating discrimination against minority communities as well. We are on the path to building our existing inequities into the digital infrastructure that will power our lives. To change that trajectory, shifting the flow of investment dollars is key. Venture Capital firms have backed nearly half of startups that graduate to the public market and have a substantial impact on innovation, particularly at the industry level. Shifting the flow of investment dollars towards equitable AI would ensure that the issue of equity in AI will be top of mind in development rooms across the globe.

Successfully shifting the flow of investment dollars towards equitable AI requires a two-pronged approach. The two-pronged approach allows lawmakers to address two core drivers of investment funding: IP and diversity. First,

185. *See* Paul Gompers & Silpa Kovvali, *The Other Diversity Dividend*, HARV. BUS. REV., July–Aug. 2018, at 72, 74 (stating that diversity significantly improves VCs' financial performance on measures such as profitable investment at the individual portfolio-company level and overall fund returns).

lawmakers should address the misaligned incentives provided by our current IP framework. Existing IP laws incentivize a lack of transparency in the AI space frustrating attempts by third parties to assess whether AI- powered products and services are inequitable. Third parties are able to observe inequitable outcomes, but are not able to drill down into the target variables or other aspects of the AI technology to determine where, if at all, bias or other kinds of inequity have leaked into its construction. Second, lawmakers should address the dearth of diversity in the investment environment. Firms' investment strategies are explicitly linked to human capital. When the demographic makeup of a VC firm is out of sync with the general population there are spillover effects into the demographic makeup of the teams those firms are likely to invest in. Because diversity of development teams is important for reducing likelihood of bias in AI, ensuring that the demographics of VC firms is aligned with the greater population is key to reducing the likelihood of inequitable AI.

There are myriad ways lawmakers could attempt to address the misaligned incentives provided by our current IP framework and the lack of diversity in the investment environment. We propose creating a fifth category of IP essentially exchanging ownership for compliance with a human rights framework and supplementing this approach with a tax incentive for VC firms graded favorably on our commitment index. For ownership under the new category of IP to attach, businesses developing and deploying AI would have to conform to five key tenets: policy commitment, active inclusion, due diligence, access to redress, and transparency. Grading on the commitment index will be publicly available and will focus on two key elements: i) does the VC firm have an optimal commitment statement with respect to diversity, and ii) has the VC taken robust steps to actualize that commitment. According to our research, optimal commitment statements should focus on long-term outcomes through internal and external actions. Whether the commitment is optimal will be evaluated under a totality of the circumstances standard, and the audit process for determining progress should focus on concrete steps taken to advance that stated commitment. Implementation of these recommendations could be the spark that helps ensure that developing equitable AI is a primary goal in every development room across the globe, leaving lasting impacts for generations to come.