



**United States House of Representatives  
Committee on Science, Space, and Technology**

**“Artificial Intelligence: Societal and Ethical Implications”  
June 26, 2019**

Written Testimony of  
Meredith Whittaker  
Co-founder and Co-director, AI Now Institute, New York University

Chairwoman Johnson, Ranking Member Lucas, and members of the Committee, thank you for inviting me to speak today. My name is Meredith Whittaker and I am the Co-founder of the AI Now Institute at New York University. AI Now is the first university research institute dedicated to studying the social implications of artificial intelligence and algorithmic technologies (“AI”).<sup>1</sup> Our work examines the rapid proliferation of AI systems through core social domains such as criminal justice, health care, employment, and education. In particular we focus on concerns in the areas of bias and inclusion, safety and critical infrastructure, rights and liberties, and labor. As we identify problems in each of these spaces, we work to address them through robust research investigations, community engagement, and key policy interventions.

When Kate Crawford and I founded AI Now in 2016, we were just beginning to see both the extreme promises of AI as well as the extreme risks. Our annual reports have chronicled these risks, issues, and concerns, and I can say with confidence that they have only increased over time.<sup>2</sup> From education, to healthcare, to law enforcement, to hiring and worker management,

---

<sup>1</sup> *About*, AI NOW INSTITUTE, <https://ainowinstitute.org/about.html/>.

<sup>2</sup> Kate Crawford et al., *THE AI NOW REPORT 2016* (Sept. 2016), [https://ainowinstitute.org/AI\\_Now\\_2016\\_Report.html](https://ainowinstitute.org/AI_Now_2016_Report.html); AI Now Inst., *THE AI NOW REPORT 2017* (2017), [https://ainowinstitute.org/AI\\_Now\\_2017\\_Report.html](https://ainowinstitute.org/AI_Now_2017_Report.html); Nicolas Suzor, Tess Van Geelen & Sarah Myers West, *Evaluating the Legitimacy of Platform Governance: A Review of Research and a Shared Research Agenda*, INTERNATIONAL COMMUNICATION GAZETTE (Feb. 2018), 80(4), pp. 385-400, <https://eprints.qut.edu.au/112749/3/112749a.pdf>; AI Now Inst., *LITIGATING ALGORITHMS: CHALLENGING GOVERNMENT USE OF ALGORITHMIC DECISION SYSTEMS* (Sept. 2018), <https://ainowinstitute.org/litigatingalgorithms.html>; Kate Crawford & Vladan Joler, *Anatomy of an AI System: The Amazon Echo As an Anatomical Map of Human Labor, Data and Planetary Resources*, AI NOW INST. & SHARE LAB (Sept. 7, 2018), <https://anatomyof.ai>; AI Now Inst., *ALGORITHMIC ACCOUNTABILITY POLICY TOOLKIT* (Oct. 2018), <https://ainowinstitute.org/aap-toolkit.pdf>; <https://ainowinstitute.org/litigatingalgorithms.pdf>; Meredith Whittaker et al., *THE AI NOW REPORT 2018* (Dec. 2018), [https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.pdf](https://ainowinstitute.org/AI_Now_2018_Report.pdf); Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 192 (April 2019), available at <https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson-Schultz-Crawford.pdf>; Sarah Myers West, Meredith Whittaker & Kate Crawford, *DISCRIMINATING SYSTEMS: GENDER, RACE AND POWER IN AI* (April 2019), <https://ainowinstitute.org/discriminatingsystems.html>.

and well beyond, the role of AI in our core social institutions is expanding, both in government and the private sector. AI is shaping access to resources and opportunity, with profound implications for hundreds of millions of Americans. These systems are being used to judge who should be released on bail;<sup>3</sup> to automate disease diagnosis in patients;<sup>4</sup> to amplify mass surveillance efforts;<sup>5</sup> and to hire, monitor and manage workers.<sup>6 7</sup> These are only a handful of examples that provide a view of the power that AI is quietly gaining over our lives and institutions.

At the same time that AI systems are proliferating and concentrating the power to impact our lives, they are slipping further away from core democratic protections such as due process and other forms of accountability. Embedded within technological and legal “black boxes,” AI systems raise many questions and provide few true answers.

In light of this, it is urgent that we address this accountability gap and ensure these technologies advance the values of fairness and justice that this institution and many others are dedicated to upholding.

### **The Emerging Problem Space**

Thanks to researchers and investigative journalists, we have significant and alarming evidence of the way in which AI systems encode bias, produce harm, accelerate environmental degradation and the climate crisis,<sup>8 9</sup> avoid established accountability processes, and differ dramatically from many of the marketing claims made by AI companies.

Take for example the simple act of listening. Many AI systems, such as Amazon’s Echo device are constantly recording every sound within range of their microphones. This is, in theory, so they can respond to user voice commands. But they are equally capable of other, unauthorized

---

<sup>3</sup> Sam Levin, *Imprisoned by Algorithms: The Dark Side of California Ending Cash Bail*, THE GUARDIAN (Sept. 7, 2018), <https://www.theguardian.com/us-news/2018/sep/07/imprisoned-by-algorithms-the-dark-side-of-california-ending-cash-bail/>.

<sup>4</sup> *Seeing Potential*, GOOGLE STORIES (2018), <https://about.google/intl/en/stories/seeingpotential/>.

<sup>5</sup> Russell Brandom, *Facial Recognition Is Coming to US Airports, Fast-Tracked by Trump*, THE VERGE (Apr. 18, 2017), <https://www.theverge.com/2017/4/18/15332742/us-border-biometric-exit-facial-recognition-scanning-homeland-security/>.

<sup>6</sup> Terena Bell, *This Bot Judges How Much You Smile During Your Job Interview*, FAST COMPANY (Jan. 15, 2019), <https://www.fastcompany.com/90284772/this-bot-judges-how-much-you-smile-during-your-job-interview/>.

<sup>7</sup> Kevin Roose, *A Machine May Not Take Your Job, but One Could Become Your Boss*, N.Y. TIMES (June 23, 2019), <https://www.nytimes.com/2019/06/23/technology/artificial-intelligence-ai-workplace.html/>.

<sup>8</sup> anatomyof.ai

<sup>9</sup> Emma Strubell, Ananya Ganesh, Andrew McCallum, *Energy and Policy Considerations for Deep Learning in NLP*, IN THE 57TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (Jun. 5 2019), <https://arxiv.org/abs/1906.02243>.

recordings.<sup>10</sup> Even when welcomed into our homes, voice recognition systems are far from objective in what they “hear” and whose voices count. For example, some systems have been shown to recognize more masculine sounding voices better than feminine voices;<sup>11</sup> such bias also exists in facial recognition systems that fail to recognize black and transgendered faces;<sup>12 13</sup> automated hiring systems that discriminate against women candidates; medical diagnostic systems don’t work for dark skinned patients;<sup>14</sup> sentencing algorithms that discriminate against black defendants; and the list goes on.

But even when these systems don’t explicitly encode bias, they can still cause harm. For example, the ACLU tested the use of Amazon’s Facial Recognition system, finding that 28 members of Congress were falsely matched with mugshots of those previously arrested for a crime.<sup>15</sup> The problems raised by the application of facial recognition and other AI systems won’t be solved by ensuring the technology is 100% accurate. The application of AI to monitor, track, and control vulnerable populations raises fundamental issues, reminding us that questions of justice will not be solved simply by adjusting a technical system.

Even in the face of mounting evidence, the rapid integration of AI into sensitive social domains continues. Government agencies are increasingly using AI and algorithmic systems to assess beneficiaries of social services and manage benefit allocation. In many cases, the outcome of these experiments has been harmful and even deadly to the people such programs are meant to serve.<sup>16</sup> For example, several states have turned to automation for Medicaid benefit allocation. In many cases, these systems have failed due to flaws in the system itself, resulting in serious harm and multi-million dollar lawsuits.<sup>17</sup> In Arkansas, such a system was used to

---

<sup>10</sup> Niraj Chokshi, *Is Alexa Listening? Amazon Echo Sent Out Recording of Couple’s Conversation*, N.Y. TIMES (May 25, 2018),

<https://www.nytimes.com/2018/05/25/business/amazon-alexa-conversation-shared-echo.html/>.

<sup>11</sup> Rachael Tatman, *Google’s Speech Recognition Has a Gender Bias*, MAKING NOISE AND HEARING THINGS (July 12, 2016),

<https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/>

<sup>12</sup> Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, GENDER SHADES (2018), [gendershades.org](http://gendershades.org).

<sup>13</sup> Jacob Snow, *Amazon’s Face Recognition Falsely Matched 28 Members of Congress With Mugshots*, ACLU (July 26, 2018),

<https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28/>.

<sup>14</sup> Angela Lashbrook, *AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind*, THE ATLANTIC (Aug. 16, 2018),

<https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/>

<sup>15</sup> See Jacob Snow, *supra* note 13.

<sup>16</sup> Virginia Eubanks, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (St. Martin’s Press 2018).

<sup>17</sup> Arkansas: Colin Lecher, *What Happens When an Algorithm Cuts Your Health Care*, THE VERGE (Mar. 21, 2018),

<https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy/>;

Idaho: *Federal Court Rules Against Idaho Department of Health and Welfare in Medicaid Class Action*, ACLU (Mar. 30, 2016),

<https://www.aclu.org/press-releases/federal-court-rules-against-idaho-department-health-and-welfare-me>

calculate how much home healthcare chronically ill Medicaid patients would receive. Due to an error, the system was significantly underprovisioning many people who required such care to survive. Patients were left to sit in their own waste, unable to access food when they were hungry, or to turn themselves to prevent bedsores.<sup>18</sup> If Legal Aid of Arkansas had not brought a case, and ultimately audited the system, it's possible that such harm would have persisted unchecked.<sup>19</sup>

It is not an overstatement to claim that the integration of AI is poised to reshape core domains. In education, we're seeing the implementation of facial recognition,<sup>20</sup> automated mental health monitoring,<sup>21</sup> and AI-based 'learning programs' that promise to track student progress and direct teachers.<sup>22</sup> These are often sold to school districts by vendors making claims about technical capabilities that, when we're able to examine them, frequently fall short. For example, a company called Gaggle licenses services to school districts that it claims increase safety and automatically detect students with mental health issues.<sup>23</sup> Gaggle's system requires blanket surveillance of student social media and private student communications sent using school networks. It uses AI to analyze the content of these communications, and to flag threats and identify students at risk. But on examination, it becomes clear that much of what gaggle detects are "minor violations," including use of profanity in private communications.<sup>24</sup> The company's claims that it helps students and improves safety are not firmly substantiated.

We are left with an understanding of Gaggle's sales pitch, but without answers to a number of socially significant questions that its technology raises: we don't know how a student might contest the company's interpretation of their communications, whether parents were fully informed before consenting to such surveillance, the way in which such monitoring may chill communication between peers, or how Gaggle's system might encode bias that could harm students with disabilities, students of color, or LGBTQ students (although we do know that in one school district the system flagged over a dozen students for communications that included

---

dicaid-class-action; Indiana: Alyssa Edes & Emma Bowman, *'Automating Inequality': Algorithms in Public Services Often Fail the Most Vulnerable*, ALL THINGS CONSIDERED (Feb. 19, 2018), <https://www.npr.org/sections/alltechconsidered/2018/02/19/586387119/automating-inequality-algorithms-in-public-services-often-fail-the-most-vulnerab/>.

<sup>18</sup> Caitlan Butler, *Changes to Medicaid Program Could Affect Union County Residents, Businesses*, EL DORADO NEWS-TIMES (Nov. 14, 2018), <https://www.eldoradonews.com/news/2018/nov/14/changes-medicaid-program-could-affect-county/>.

<sup>19</sup> See Colin Lecher, *supra* note 17.

<sup>20</sup> Mariella Moon, *Facial recognition is coming to US schools, starting in New York*, ENGADGET (May 30, 2019), <https://www.engadget.com/2019/05/30/facial-recognition-us-schools-new-york/>.

<sup>21</sup> Kaveh Waddell, *Schools Turn to AI to Monitor Students' Mental Health*, AXIOS (August 29, 2018), <https://www.axios.com/school-filtering-for-self-harm-prevention-4bf481cc-a351-4f59-8cb3-318943548edc.html>

<sup>22</sup> <https://www.century.tech/>

<sup>23</sup> <https://www.gaggle.net/>

<sup>24</sup> Benjamin Herold, *Schools Are Deploying Massive Digital Surveillance Systems. The Results Are Alarming*, EDUCATION WEEK (May 30, 2019), <https://www.edweek.org/ew/articles/2019/05/30/schools-are-deploying-massive-digital-surveillance-systems.html>

the word “gay”).<sup>25</sup> Gaggle is one example among many in which AI is being applied in ways that have profound and life-changing implications without first ensuring that it is safe, helpful, unbiased, and doesn't pose risks to those it's meant to serve.

## Corporate Control and Secrecy

AI technology affecting the lives and opportunities of hundreds of millions of people is being shaped by the incentives, values, and interests of a small handful of private companies.

AI is not a new set of technologies, and many of the core techniques that power AI systems, including neural nets, have been around for decades. The biggest changes recently have not been wholly new AI techniques (although we have seen improvements and innovations). What has changed drastically is the availability of massive amounts of data and vast computational resources. It's this that is behind the AI boom we see today. These are assets that only a handful of major tech companies have, and very few others do. This is one of the reasons why the US government contracts with companies like Amazon, instead of building its own infrastructure and AI. Without the legal protections afforded the private sector around privacy, and existing market reach and infrastructural economies of scale, it's virtually impossible to obtain the resources needed to create AI from scratch.

While there are many AI startups, most have significant budget allocated to licensing computational infrastructure from one of the big tech companies – namely Amazon, Microsoft, or Google. Most also struggle to acquire data, often scraping the web, buying data from data brokers, or signing deals in one or another domain (say, healthcare or education) to get access to relevant datasets. In short, the startup AI ecosystem we see today ultimately relies on the infrastructure provided by big companies, and thus the AI industry points back to a few large players.

This means that much of what we know about AI is written by the public relations and marketing departments of these same companies. They highlight benevolent uses and public benefit, often showcasing prototypes that haven't been validated beyond narrow test cases,<sup>26</sup> while remaining silent about the application of AI to fossil fuel extraction,<sup>27</sup> weapons development,<sup>28</sup> mass

---

<sup>25</sup> See *id.*

<sup>26</sup> Yossi Matias, *Keeping People Safe With AI-Enabled Flood Forecasting*, THE KEYWORD (Sep. 24, 2018), <https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/>.

<sup>27</sup> Brian Merchant, *How Google, Microsoft, and Big Tech Are Automating the Climate Crisis*, GIZMODO (Feb. 21, 2019), <https://gizmodo.com/how-google-microsoft-and-big-tech-are-automating-the-1832790799/>.

<sup>28</sup> Jason Evangelho, *Microsoft Employees Upset About HoloLens As U.S. Military Weapon*, FORBES (Feb. 23, 2019), <https://www.forbes.com/sites/jasonevangelho/2019/02/23/microsoft-employees-upset-about-hololens-as-u-s-military-weapon/#5e49f1524822/>.

surveillance,<sup>29</sup> or the problems of bias and error.<sup>30</sup> We know about many of these more troubling cases thanks to researchers, investigative journalists, and whistleblowers.

When regulators, researchers, and the public seek to learn more, and to research and understand the potential harms of these systems, they are faced with structural barriers. The companies developing and deploying AI often exploit corporate secrecy laws, making testing, auditing, and monitoring extremely difficult, if not impossible.<sup>31</sup>

This structural secrecy, combined with overbroad laws used to threaten scientists, journalists, and algorithmic auditors,<sup>32</sup> makes it very difficult to validate company marketing promises. Access to fundamental information about AI systems, like where, how, and to what end they're being used, are classed as proprietary and confidential. Often even workers within these firms don't know where, and how, technology they contribute to will ultimately be applied.<sup>33</sup>

### **Lack of Accountability to Those Most at Risk**

This pattern is particularly concerning given that those who will be most affected by such systems are rarely part of decision-making leading to the purchase and implementation of such systems, and often lack power to challenge subsequent uses. Tenants in rent-stabilized buildings in Brooklyn only learned of their landlord's plan to install a StoneLock brand facial recognition entry system after he made the decision to procure and install the system. A letter from New York State's Homes and Community Renewal (HCR), the agency overseeing rent-stabilized housing, informed tenants of the landlord's application to install the system. The tenants pushed back, citing well-documented problems of bias and inaccuracy, as well as the privacy concerns inherent in providing sensitive biometric data to a landlord with a history of punitive tenant surveillance.<sup>34</sup> Currently, the plan remains in limbo, with a lawyer representing the tenants noting that HCR does not have the jurisdiction or authority to adequately protect

---

<sup>29</sup> Drew Harwell, *Oregon Became a Testing Ground for Amazon's Facial-Recognition Policing. But What If Rekognition Gets It Wrong?*, WASHINGTON POST (Apr. 30, 2019), [https://www.washingtonpost.com/technology/2019/04/30/amazons-facial-recognition-technology-is-supercarging-local-police/?utm\\_term=.50977aa76349/](https://www.washingtonpost.com/technology/2019/04/30/amazons-facial-recognition-technology-is-supercarging-local-police/?utm_term=.50977aa76349/).

<sup>30</sup> Jeremy B. Merrill & Ariana Tobin, *Facebook Moves to Block Ad Transparency Tools — Including Ours*, PROPUBLICA (Jan. 28, 2019), <https://www.propublica.org/article/facebook-blocks-ad-transparency-tools/>.

<sup>31</sup> Vinod Iyengar, *Why AI Consolidation Will Create the Worst Monopoly in US History*, TECHCRUNCH (Aug. 24, 2016), <https://techcrunch.com/2016/08/24/why-ai-consolidation-will-create-the-worst-monopoly-in-us-history/>.

<sup>32</sup> *Sandvig v. Barr — Challenge to CFAA Prohibition on Uncovering Racial Discrimination Online*, ACLU (May 22, 2019), <https://www.aclu.org/cases/sandvig-v-barr-challenge-cfaa-prohibition-uncovering-racial-discrimination-online/>.

<sup>33</sup> Kate Conger and Cade Metz, *Tech Workers Now Want to Know: What Are We Building This For?*, N.Y. TIMES (Oct. 7, 2018), <https://www.nytimes.com/2018/10/07/technology/tech-workers-ask-censorship-surveillance.html>

<sup>34</sup> Rashida Richardson, *Letter in Support of Brooklyn Legal Services' Opposition Re: New York State Homes and Community Renewal Docket Nos. GS210005OD and GS210008OD*, AI NOW INSTITUTE (May 1, 2019), <https://ainowinstitute.org/dhcr-amici-letter-043019.pdf>.

tenants, and encouraging the agency to reject the landlord’s application given the lack of meaningful protections.<sup>35</sup>

Agreements between a vendor selling such a system on the one hand, and a business or institution with an interest in using it on the other, are most often reached behind closed doors. In the case of many large tech companies, the fact that such an agreement has been reached may itself be considered confidential. If the Brooklyn buildings had not been rent-stabilized, and thus had not required an application to a state agency to implement such changes, it’s possible that the tenants would only have learned after a contract between the landlord and the vendor had been signed. In most cases, there’s no requirement that such agreements take the interests of people who will be tracked, classified, and surveilled by AI systems into account. And since these systems are often integrated in ways that aren’t visible – as part of larger infrastructures – those affected by them may not know that an AI system had a role in making a determination that impacted their life.<sup>36</sup>

### **Lack of Diversity and Its Implications**

The technology affecting the lives and opportunities of billions of people is being shaped by the incentives, values, and interests of a small handful of private companies. And these companies are notoriously non-diverse. AI Now conducted a year-long study on the AI industry’s diversity and inclusion practices, and the results are bleak. Women make up 10% of research staff at Google and just 15% at Facebook,<sup>37</sup> and the picture is worse when you examine available data on non-white workers. Only 2.5% of Google’s employees are black, and only 3.6% are latinx, while Facebook and Microsoft are each at 4% for black workers and 6% for latinx workers.<sup>38</sup> We have no data on trans workers and other gender minorities, but anecdotal evidence is grim. This extends beyond industry. In academia, over 80% of AI professors are men, and only 18% of authors at leading AI conferences are women.<sup>39</sup> To give an example of how stark the diversity problem is, when BlackInAI co-founder, Timnit Gebru, first attended the major AI conference NeurIPS in 2016, she counted 6 black attendees out of 8,500 total.<sup>40</sup>

---

<sup>35</sup> *Brooklyn Tenants File Legal Opposition to Landlord’s Application to Install Facial Recognition Entry System in Building*, LEGAL SERVICES NYC (May 1, 2019), <https://www.legalservicesnyc.org/news-and-events/press-room/1466-brooklyn-tenants-file-legal-opposition-to-landlords-application-to-install-facial-recognition-entry-system-in-building>

<sup>36</sup> Shannon Liao, *Chinese Facial Recognition System Mistakes a Face on a Bus for a Jaywalker*, THE VERGE (Nov. 22, 2018), <https://www.theverge.com/2018/11/22/18107885/china-facial-recognition-mistaken-jaywalker/>.

<sup>37</sup> Tom Simonite, *AI Is the Future—but Where Are the Women*, WIRED (Aug. 17, 2018), <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>.

<sup>38</sup> *Diversity & Inclusion*, MICROSOFT GLOBAL DIVERSITY AND INCLUSION, <https://www.microsoft.com/en-us/diversity/default.aspx> ; Maxine Williams, *Facebook 2018 Diversity Report: Reflecting on Our Journey*, FACEBOOK NEWSROOM (July 12, 2018), <https://newsroom.fb.com/news/2018/07/diversity-report/>.

<sup>39</sup> JF Gagne, *Global AI Talent Report 2019*, <https://jfgagne.ai/talent-2019/>

<sup>40</sup> Jackie Snow, *“We’re in a Diversity Crisis”: Cofounder of Black in AI on What’s Poisoning Algorithms in Our Lives*, MIT TECHNOLOGY REVIEW (Feb. 14, 2018),

The diversity crisis in the AI industry means that women, people of color, gender minorities, and other marginalized populations are excluded from contributing to the design of AI systems, from shaping how these systems function, and from determining what problems these systems are tasked with solving.<sup>41</sup> This influences which AI products get built, who they are designed to serve, and who benefits from their development. And in the case of AI, these determinations affect billions of people beyond company walls.

AI systems have evidenced a persistent pattern of gender and race-based discrimination (among other attributes and forms of identity), and in most cases, such bias mirrors and replicates existing structures of inequality.<sup>42</sup> To review a few examples: sentencing algorithms discriminate against black defendants;<sup>43</sup> chatbots easily adopt racist and misogynistic language when trained on online discourse;<sup>44</sup> and Uber's facial recognition doesn't work for trans drivers, locking them out of work.<sup>45</sup> Here we see a common theme: when these systems fail, they fail in ways that harm those who are already marginalized. I have yet to encounter an AI system that was biased against white men as a standalone identity.

Such bias can be the result of faulty training data, problems in how the system was designed or configured,<sup>46</sup> or bad or biased applications in real world contexts. In all cases it signals that the environments where a given system was created and envisioned didn't recognize or reflect on the contexts within which these systems would be deployed. Or, that those creating and maintaining these systems did not have the experience or background to understand the diverse environments and identities that would be impacted by a given system. Recent research from AI Now's Rashida Richardson shows vendors selling predictive policing systems failed to account for potential biases in the data such systems relied on, and thus to prevent harm.<sup>47</sup> In at least nine jurisdictions, predictive policing tools were being used or developed on data that was generated during periods where the police departments engaged in corrupt, racially biased, or

---

<https://www.technologyreview.com/s/610192/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our/>

<sup>41</sup> Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016),

<https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>

<sup>42</sup> See, e.g., Safiya Umoja Noble, *ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM* (NYU Press, 2013); Latonya Sweeney, *Discrimination in Online Ad Delivery*, 56 COMM. OF THE ACM 5, 44-45 (2013); Dorothy E. Roberts, *Book Review: Digitizing the Carceral State*, 1696 Harv. L. Rev. 1695 (2019); Muhammad Ali et al., *Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*, ARXIV (Apr. 19, 2019), <https://arxiv.org/pdf/1904.02095.pdf>.

<sup>43</sup> Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016),

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

<sup>44</sup> James Vincent, *Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day*, THE VERGE (Mar. 24, 2016), <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist/>.

<sup>45</sup> Steven Melendez, *Uber Driver Troubles Raise Concerns About Transgender Face Recognition*, FAST COMPANY (Aug. 9, 2018),

<https://www.fastcompany.com/90216258/uber-face-recognition-tool-has-locked-out-some-transgender-drivers/>.

<sup>46</sup> See Colin Lecher, *supra* note 17.

<sup>47</sup> See Rashida Richardson et al., *supra* note 2.



unlawful policing practices and policies. Police data reflects the local practices, policies, and environment where it was collected, as there are no national standards on how police data is collected or used. This research showed that reliance on such “dirty data” increases the risk that these police technologies will reproduce and perpetuate past discrimination..This is another example of how the failure by vendors and developers to account for history and context risks encoding bias directly into their systems. And thus risk the lives and futures of those wrongfully profiled and targeted.

## **Beyond Technical Solutions**

Both within the spaces where AI is being created, and in the logic of how AI systems are designed, the costs of bias, harassment, and discrimination are borne by the same people: gender minorities, people of color, and other under-represented groups. Similarly, the benefits of such systems – from profit to efficiency – accrue primarily to those already in positions of power, who again tend to be white, educated, and male.<sup>48</sup>

This points to a problem that goes beyond technical fixes for issues of AI bias and discrimination. AI systems are powerful. They allow those who use them to profoundly influence people’s lives, across multiple domains. And the ability to access and use these systems is not evenly distributed. Given the expense of creating AI, the cost of maintaining these systems, and the market incentives driving their development, it is almost always those who already have power who are in the position of applying AI systems, often on those who don’t. Immigration and Customs Enforcement (ICE) had used a risk assessment algorithm at the border since 2013, meant to help determine whether an immigrant should be detained or released on bond. An academic study examining the algorithm found that it had been modified a number of times in an attempt to mitigate the overly-punitive behavior of ICE agents,<sup>49</sup> but this technical bias fix failed to solve the problem. Then, in 2017 ICE modified the algorithm again, to *only* suggest detain, and never release. The impact of this change was profound. ICE detained over 43,000 immigrants with no criminal history in 2017, more than three times as many as the previous year.<sup>50</sup>

Here we see the way in which the problems of AI go beyond bias and inaccuracy at the technical level, to fundamental issues of power and control. This means our approach to these problems must expand beyond the technical, asking who is harmed by these systems? Who benefits? And who gets to decide?

---

<sup>48</sup> See Sarah Myers West et al., *supra* note 2.

<sup>49</sup> Robert Koulisch & Ernesto F. Calvo, *The Human Factor: Algorithms, Dissenters, and Detention in Immigration Enforcement*, ILCSS WORKING PAPER | NO. 1 (March 16, 2019), available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3355663](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3355663).

<sup>50</sup> Mica Rosenberg & Reade Levinson, *Trump’s Catch-And-Detain Policy Snares Many Who Have Long Called U.S. Home*, REUTERS (June 20, 2018), <https://www.reuters.com/investigates/special-report/usa-immigration-court/>

## Ethics Is Not Enough

In the face of evidence of bias, error, and misuse of AI systems, we have seen tech companies and the AI field beyond adopt ethical principles and guidelines, and form ethical oversight committees and boards. It is encouraging that those creating such powerful technologies are recognizing their potential harms and moving to address these.

However, these codes, guidelines and ethics boards do not substitute for meaningful accountability and oversight. To date we have no examples where such ethical proclamations are backed by public enforcement mechanisms, nor any that attach clear consequences for failing to live up to ethical ideals. While the public may be able to compare a decision already made by an AI company to its guiding principles, this allows no insight into decision-making, nor does it give anyone outside of the company the power to reverse or guide such a decision.

Such a turn to ethics may also serve to stall movements toward more substantive regulation and public accountability measures, working to deflect criticism by acknowledging that problems exist, without ceding any power to regulate or transform the way technology is developed and applied.<sup>51</sup>

The case of Google is instructive here. Among the list of applications that Google promises not to pursue as a part of its AI Principles are “technologies whose purpose contravenes widely accepted principles of international law and human rights.”<sup>52</sup> This commitment was tested in 2018 when whistleblowers revealed that the company was quietly developing a censored version of its search engine (which relies extensively on AI capabilities) for the Chinese market, code-named Dragonfly.<sup>53</sup> Human rights organizations condemned the project as a violation of human rights law, and as such, a violation of Google’s AI principles.<sup>54 55</sup> While there are indications that the project has been cancelled following Google worker organizing and public outcry, its continued development in the face of the company’s principles was never explained,

---

<sup>51</sup> Ben Wagner, *Ethics as Escape From Regulation: From Ethics-Washing to Ethics-Shopping?*, in BEING PROFILED: COGITAS ERGO SUM, ed. Mireille Hildebrandt. (Amsterdam University Press, forthcoming 2019), [https://www.privacylab.at/wp-content/uploads/2018/07/Ben\\_Wagner\\_Ethics-as-an-Escape-from-Regulation\\_2018\\_BW9.pdf](https://www.privacylab.at/wp-content/uploads/2018/07/Ben_Wagner_Ethics-as-an-Escape-from-Regulation_2018_BW9.pdf).

<sup>52</sup> Sundar Pichai, *AI at Google: Our Principles*, THE KEYWORD, June 7, 2018, <https://www.blog.google/technology/ai/ai-principles/>.

<sup>53</sup> Ryan Gallagher, *Google Plans to Launch Censored Search Engine in China, Leaked Documents Reveal*, THE INTERCEPT, August 1, 2018, <https://theintercept.com/2018/08/01/google-china-search-engine-censorship/>.

<sup>54</sup> Ronald Deibert, Rebecca Mackinnon, Xiao Qiang, and Lokman Tsui, *Open Letter to Google on Reported Plans to Launch a Censored Search Engine in China*, AMNESTY INTERNATIONAL (August 28, 2018), <https://www.amnesty.org/en/documents/document/?indexNumber=ASA17%2f9001%2f2018>.

<sup>55</sup> Kate Conger and Daisuke Wakabayashi, *Google Employees Protest Secret Work on Censored Search Engine for China*, N.Y. TIMES, September 10, 2018, <https://www.nytimes.com/2018/08/16/technology/google-employees-protest-search-censored-china.html>.

and the principles themselves seem to have had little impact on directing the company to reconsider the effort.<sup>56</sup>

## Organizing and Accountability

Over the past year, we have seen mounting pushback, as organizing and resistance to the development and application of AI systems has grown. Amazon Warehouse workers in Minneapolis protested Amazon's automated management system, pushing back against a system of worker control that continually extracted more labor at the expense of workers' health and well-being.<sup>57</sup> Similarly, Uber drivers organized a nationwide strike, protesting the centralized, algorithmic platform that arbitrarily cuts driver wages without recourse.<sup>58</sup> And even students at a school where Facebook-backed "personalized" education systems were used staged protests, demanding less dehumanizing forms of education.<sup>59</sup>

We've also seen pushback inside tech companies. Workers across the industry have organized and protested unethical and secretive projects,<sup>60 61 62 63 64</sup> demanding to have a say in what they

---

<sup>56</sup> Mark Bergen, *Google CEO Tells Staff China Plans Are 'Exploratory' After Backlash*, BLOOMBERG (August 17, 2018),

<https://www.bloomberg.com/news/articles/2018-08-17/google-ceo-is-said-to-tell-staff-china-plans-a-re-exploratory>

<sup>57</sup> Chavie Lieber, *Muslim Amazon Workers Say They Don't Have Enough Time to Pray. Now They're Fighting for Their Rights*, VOX (Dec. 17, 2018),

<https://www.vox.com/the-goods/2018/12/14/18141291/amazon-fulfillment-center-east-africa-workers-minneapolis>

<sup>58</sup> Faiz Siddiqui, *Uber and Lyft Drivers Strike for Pay Transparency — After Algorithms Made It Harder to Understand*, WASHINGTON POST (May 8, 2019),

[https://www.washingtonpost.com/technology/2019/05/08/uber-lyft-drivers-strike-pay-transparency-after-algorithms-made-it-harder-understand/?utm\\_term=.b4c939121e0e](https://www.washingtonpost.com/technology/2019/05/08/uber-lyft-drivers-strike-pay-transparency-after-algorithms-made-it-harder-understand/?utm_term=.b4c939121e0e)

<sup>59</sup> Susan Edelman, *Brooklyn Students Hold Walkout in Protest of Facebook-Designed Online Program*, N.Y. POST (Nov. 10, 2018),

<https://nypost.com/2018/11/10/brooklyn-students-hold-walkout-in-protest-of-facebook-designed-online-program/>

<sup>60</sup> Dave Lee, *Microsoft Staff: Do Not Use HoloLens for War*, BBC NEWS (Feb. 22, 2019),

<https://www.bbc.com/news/technology-47339774/>.

<sup>61</sup> Ryan Gallagher, *Google Dragonfly*, THE INTERCEPT (Aug. 1, 2018 – May 14, 2019),

<https://theintercept.com/collections/google-dragonfly-china/> (collecting 25 articles of Google Dragonfly coverage).

<sup>62</sup> Kate Conger, *Amazon Workers Demand Jeff Bezos Cancel Face Recognition Contracts With Law Enforcement*, GIZMODO (June 21, 2018),

<https://gizmodo.com/amazon-workers-demand-jeff-bezos-cancel-face-recognition-1827037509>

<sup>63</sup> Scott Shane and Daisuke Wakabayashi, *'The Business of War': Google Employees Protest Work for the Pentagon*, N.Y. TIMES (April 4, 2018),

<https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html/>.

<sup>64</sup> Becky Peterson, *Salesforce Employees Are Upset Over the Company's Work With U.S. Customs and Border Protection as Silicon Valley Grapples With the Government's Use of Tech*, BUSINESS INSIDER PRIME (Jun 25, 2018),

<https://www.businessinsider.com/salesforce-employees-protest-work-customs-border-protection-agency-2018-6/>.

build,<sup>65</sup> and connecting AI's ethical issues with problems of workplace discrimination, harassment, and abuse.<sup>66</sup>

Currently, such organizing and activism comprises one of the primary modes of accountability working to check the biased and oppressive deployment of AI technologies, and to improve diversity and equity across the industry.<sup>67 68 69 70</sup> I have proudly organized with fellow tech workers at Google and beyond, recognizing the urgent need for such accountability.<sup>71</sup> What is clear to those of us who have been researching these issues, and those of us familiar with the tech industry, is that we need much more. Accountability, transparency, oversight, and measures that ensure those most at risk of harm are at the heart of AI decision making.<sup>72 73</sup>

### **The Path Forward: AI Now's Policy Recommendations for Congress:**

Congress has a window to act, and the time is now. Powerful AI systems are currently reshaping our lives and social institutions in ways that we aren't able to measure or contest. These systems are developed and deployed by a handful of private companies whose market interests don't always align with the public good, and who shield these systems from accountability behind claims of corporate secrecy. When we are able to examine these systems, too often we find that they are biased and inaccurate in ways that replicate historical patterns of discrimination.

It is imperative that lawmakers regulate to ensure that these systems are accountable, accurate, contestable, and that those most at risk of harm have a say in how, and whether, they are used. As first steps toward this goal, AI Now recommends that lawmakers:

---

<sup>65</sup> See Kate Conger and Daisuke Wakabayashi, *supra* note 55.

<https://www.nytimes.com/2018/08/16/technology/google-employees-protest-search-censored-china.html/>.

<sup>66</sup> Google Walkout for Real Change, *20,000 Google Employees and Contractors Participate in Global "Walkout For Real Change"* (Nov 2, 2018),

<https://medium.com/@GoogleWalkout/google-employees-and-contractors-participate-in-global-walkout-for-real-change-389c65517843/>.

<sup>67</sup> Daisuke Wakabayashi, *Google Ends Forced Arbitration for All Employee Disputes*, N.Y. TIMES (Feb. 21, 2019), <https://www.nytimes.com/2019/02/21/technology/google-forced-arbitration.html/>.

<sup>68</sup> BBC News, *Microsoft Responds to Female Harassment Claims*, BBC (April 5, 2019), <https://www.bbc.com/news/technology-47826249/>.

<sup>69</sup> Nick Bastone, *Google Is Likely to End Its Efforts to Build a Censored Search Engine for China, Says Report*, BUSINESS INSIDER (Dec. 17, 2018),

<https://www.businessinsider.com/google-dragonfly-china-canceled-2018-12/>.

<sup>70</sup> Daisuke Wakabayashi & Scott Shane, *Google Will Not Renew Pentagon Contract that Upset Employees*, N.Y. TIMES (June 1, 2018),

<https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html/>.

<sup>71</sup> Kyle Wiggers, *How Google Treats Meredith Whittaker Is Important to Potential AI Whistleblowers*, VENTUREBEAT (April 24, 2019),

<https://venturebeat.com/2019/04/24/how-google-treats-meredith-whittaker-is-important-to-potential-ai-whistleblowers/>.

<sup>72</sup> Frank Pasquale, *Odd Numbers: Algorithms Alone Can't Meaningfully Hold Other Algorithms Accountable*, Real Life Mag (Aug. 20, 2018), <https://reallifemag.com/odd-numbers/>.

<sup>73</sup> See Meredith Whittaker et al., *supra* note 2.

## **1. Require Algorithmic Impact Assessments in both Public and Private Sectors before AI Systems are Acquired and Used**

In 2018, AI Now published an Algorithmic Impact Assessment (AIA) framework, which offers a means for assessing algorithmic systems in government, including AI based systems, and providing the public with transparency and decision-making power.<sup>74</sup> AIAs draw directly from impact assessment frameworks in environmental protection, human rights, privacy, and data protection policy domains by combining public agency review and public input.<sup>75</sup> When implemented in government, AIAs provide both the agency and the public the opportunity to evaluate the potential impacts of the adoption of an algorithmic system before the agency has committed to its use. AIAs also require ongoing monitoring and review, recognizing the dynamic contexts within which such systems are applied.

The framework has been adopted in Canada, and is being considered by local, state, and national governments globally.<sup>76</sup> Though it was originally proposed to address concerns associated with government use of automated decision-making systems, the framework should also be mandated for private companies, ensuring review and public engagement before a product or service is used in ways that impact the public. This can provide companies opportunities to assess and possibly mitigate adverse or unanticipated outcomes during the development process. It also provides the government and public with greater transparency, along with a means to strengthen existing consumer accountability mechanisms. By requiring that proposed technologies be reviewed by the communities who will be most affected by their use, AIAs work to empower those most at risk, and to encourage the development of safer and more ethical technologies.

## **2. Require Technology Companies to Waive Trade Secrecy and Other Legal Claims That Hinder Oversight and Accountability Mechanisms**

Corporate secrecy laws are a barrier to oversight, accountability, and due process when they are used to obscure technologies used in ways that affect the public. They can

---

<sup>74</sup> See AI NOW INST., ALGORITHMIC ACCOUNTABILITY POLICY TOOLKIT, *supra* note 2.

<sup>75</sup> Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

<sup>76</sup> EUROPEAN PARLIAMENT PANEL FOR THE FUTURE OF SCI. AND TECH., A GOVERNANCE FRAMEWORK FOR ALGORITHMIC ACCOUNTABILITY AND TRANSPARENCY: STUDY (Apr. 4, 2019), [http://www.europarl.europa.eu/stoa/en/document/EPRS\\_STU\(2019\)624262](http://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2019)624262); Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong., (1st Sess. 2019), <https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf>; Canada Treasury Board's Directed Automated Decision-Making, AUTONOMISATION DES ACTEURS JUDICIAIRES PAR LA CYBERJUSTICE (Nov. 25, 2018), <https://www.ajcact.org/2018/11/25/canada-treasury-boards-directive-on-automated-decision-making/>

inhibit necessary government oversight and enforcement of consumer protection laws,<sup>77</sup> which contribute to the “black box effect,” making it hard to assess bias, contest decisions, or remedy errors. Anyone procuring these technologies for use in the public sector should have the right to demand vendors waive these claims before entering into any agreements. Additionally, limiting the use of these legal claims across the board will help facilitate better oversight by state and federal consumer protection agencies and enforcement of false and deceptive practice laws.

### **3. Require Public Disclosure of Technologies That Are Involved in Any Decisions About Consumers by Name and Vendor/s**

Those most affected by AI’s use in sensitive social domains often don’t know that a given system was used, or where and how it might have shaped their lives. Consumers deserve to know about which AI systems are making decisions about them, or affecting the types of services, resources, or opportunities made available to them. Requiring disclosure of which AI systems are used, in what context, along with which companies developed such systems, will provide consumers with the notice necessary to enforce their due process rights. We need to ensure meaningful insight and transparency. This is especially urgent given that infrastructure owned by the major technology companies is often licensed and used by other businesses. Large technology companies license AI application program interfaces (APIs), or “AI as a service” to third parties, who apply them to one or another purpose.<sup>78</sup> These business relationships, in which one organization repurposes potentially flawed and biased AI systems created by large technology companies, are rarely disclosed to the public, and even the fact that there is such a relationship is often protected under nondisclosure agreements. It is hard, if not impossible, to simply know if an AI model is being used in a given context, let alone to know that such a model was created by Facebook, Google, or Amazon. Thus, understanding the implications of bad, biased, or misused models is effectively impossible, especially for those most at risk.

### **4. Enhance Whistleblower Protections and Protections for Conscientious Objectors within Technology Companies**

Organizing and resistance by tech workers has emerged as a force for accountability

---

<sup>77</sup> *Houston Fed’n of Teachers, Local 2415 v. Houston Indep. Sch. Dist.*, 251 F.Supp.3d 1168 (S.D. Tex. 2017).

<sup>78</sup> *Cognitive Services*, MICROSOFT AZURE, <https://azure.microsoft.com/en-us/services/cognitive-services/> (last visited June 16, 2019); *AI and Machine Learning Products*, GOOGLE CLOUD, <https://cloud.google.com/products/ai/> (last visited June 16, 2019); *Tools for Advancing the World’s AI*, FACEBOOK ARTIFICIAL INTELLIGENCE, <https://ai.facebook.com/tools/> (last visited June 16, 2019); *Machine Learning on Amazon AWS*, AMAZON WEB SERVICES, <https://aws.amazon.com/machine-learning/> (last visited June 16, 2019); Matt Murphy & Steve Sloane, *The Rise of APIs*, TECHCRUNCH (May 21, 2016), <https://techcrunch.com/2016/05/21/the-rise-of-apis/>.

and ethical oversight.<sup>79</sup> Alongside organizers, whistleblowers have provided a crucial public benefit, revealing products and problems that may not otherwise be visible to relevant oversight bodies, or even to the majority of workers within a given company. Whistleblowers in the technology industry can be a crucial component to government oversight, serving the public interest by revealing troubling and unethical initiatives poised to affect millions of people. In light of their service, and the critical role they are playing, they should be provided enhanced protections.

---

<sup>79</sup> Daisuke Wakabayashi & Scott Shane, *Google Will Not Renew Pentagon Contract that Upset Employees*, N.Y. TIMES (June 1, 2018), <https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html>; Avie Schneider, *Microsoft Workers Protest Army Contract With Tech 'Designed to Help People Kill'*, NPR (Feb. 22, 2019), <https://www.npr.org/2019/02/22/697110641/microsoft-workers-protest-army-contract-with-tech-designed-to-help-people-kill>; Mark Bergen & Nico Grant, *Salesforce Staff Ask CEO to Revisit Ties with Border Agency*, BLOOMBERG (June 25, 2018), <https://www.bloomberg.com/news/articles/2018-06-25/salesforce-employees-ask-ceo-to-revisit-ties-with-border-agency>.