TESTIMONY OF NICOL TURNER LEE
Fellow, Center for Technology Innovation, Brookings Institution

Before the
Task Force on Artificial Intelligence
United States House Committee on Financial Services

Hearing on "Perspectives on Artificial Intelligence: Where We Are and the Next Frontier in Financial Services"

June 26, 2019

Chairwoman Waters, Ranking Member McHenry and Members of the Committee, thank you for the opportunity to testify. I am encouraged by the interest of this committee on artificial intelligence (AI) and the application of autonomous systems to the financial services sector. I am Nicol Turner Lee, Fellow in the Center for Technology Innovation at the Brookings Institution. With a history of over 100 years, Brookings is committed to evidenced-based, nonpartisan research in a range of focus areas. My particular research expertise encompasses data collection and analysis around regulatory and legislative policies that govern telecommunications and high-tech companies, along with the impacts of digital exclusion, artificial intelligence and machine-learning algorithms on vulnerable consumers. My forthcoming book, *The digitally invisible: How the internet is creating the new underclass,* also addresses this topic.

Introduction

Increasingly, the private and public sectors are turning to artificial intelligence (AI) systems and machine learning algorithms to automate simple and complex decision-making processes. The mass-scale digitization of data and the emerging technologies that use them are disrupting most economic sectors, including transportation, retail, advertising, financial services and energy, and other areas. AI is also having an impact on democracy and governance as computerized systems are being deployed to improve accuracy and drive objectivity in government functions.

It is the availability of massive data sets which has made it easy to derive new insights through computers. As a result, *machine learning algorithm*s, which are a set of step-by-step instructions that computers follow to perform a task, have become more sophisticated and pervasive tools for automated decision-making.[1] While algorithms are used in many contexts from making recommendations about movies to credit products, I rely on a definition that I made recently in a newly released paper[2] which refers to them as computer models that make inferences from data about people, including their identities, their demographic attributes, their preferences, and their likely future behaviors, as well as the objects related to them.[3]

In machine learning, algorithms rely on multiple data sets, or training data, that specifies what the correct outputs are for some people or objects. From that training data, it then learns a model which can be applied to other people or objects and make predictions about what the correct outputs should be for them.[4] However, because machines can treat similarly-situated people and objects differently, research is starting to reveal some troubling examples in which the reality of algorithmic decision-making falls short of our expectations, or is simply wrong. For example, automated risk assessments used by U.S. judges to determine bail and sentencing limits can generate incorrect conclusions, resulting

---

[1] The concepts of AI, algorithms and machine learning are often conflated and used interchangeably. In this paper, we will follow generally understood definitions of these terms as set out in publications for the general reader. See, e.g., Stephen F. DeAngelius. "Artificial intelligence: How algorithms make systems smart," Wired Magazine, September 2014. Available at https://www.wired.com//insights/2014/09/artificial-intelligence-algorithms-2/ (last accessed June 25, 2019). See also, Michael J. Garbade. "Clearing the Confusion: AI vs. Machine Learning vs. Deep Learning Differences," Towards Data Science, September 14, 2018. Available at https://towardsdatascience//clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb (last accessed April 12, 2019).

[2] Turner Lee, N., Resnick, P., and Barton, G. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings. (2019). Available at https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ (last accessed June 25, 2019). Angwin, J. Tobin, A. Varner, M. (2017). Facebook (Still) Letting Housing Advertisers Exclude Users by Race. *ProPublica.* Available at: https://goo.gl/Vk4jrs (last accessed June 25, 2019).

[3] Andrea Blass and Yuri Gurevich. Algorithms: A Quest for Absolute Definitions. Bulletin of European Association for Theoretical Computer Science 81, 2003. https://www.microsoft.com/en-us/research/wp-content/uploads/2017/01/164.pdf (last accessed June 25, 2019).

[4] Technically, this describes what is called "supervised machine learning."

in large cumulative effects on certain groups, like longer prison sentences or higher bails imposed on people of color. Or, credit decisions based on inferential data about applicants, such as their zip code, social media profiles or web browsing histories, can lead to higher rejection rates.

Referring back to my recent paper on this subject, my co-authors and I determine that an algorithmic decision generates "bias" when its outcomes are systematically less favorable to individuals within a particular group and where there is no relevant difference between groups that justifies such harms.[5] Bias in algorithms can come from unrepresentative or incomplete training data, or the reliance on flawed information that reflects historical inequalities. The bottom line is that if left unchecked, biased algorithms can lead to decisions which can have a collective, disparate impact on certain groups of people even without the programmer's intention to discriminate.

In my testimony, I hope to further unpack the concept of algorithmic bias and outline why we need to proactively work to identify and mitigate online biases. I conclude this written testimony with a series of recommendations – whether driven by policymakers or the self-regulatory actions of industries - that can facilitate more ethical, fair and just algorithmic models. If not carefully identified and mitigated, algorithms – especially those associated with the sensitive use cases to be discussed in this hearing – have the potential to replicate and amplify stereotypes historically prescribed to people of color and other vulnerable populations.

Racial and Ethnic Biases in the Online Economy

I'd like to start with an *initial truth* about emerging technologies. Despite their facilitation of greater efficiencies and cognition due to the programming of machines, the online economy has not resolved

[5] Blog. "Understanding bias in algorithmic design," Impact.Engineered, September 5, 2017. Available at https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e (last accessed June 25, 2019). This definition is intended to include the concepts of disparate treatment and disparate impact, but the legal definitions were not designed with AI in mind. For example, the demonstration of disparate treatment does not describe the ways in which an algorithm can learn to treat similarly situated groups differently, as will be discussed later in the paper.

the issue of racial bias in its applications. In 2013, online search results for "black-sounding" names were more likely to link arrest records with profiles, even when false.[6] Two years later, Google apologized for an algorithm that automatically tagged and labeled two African Americans as "gorillas" after an innocuous online word search.[7] In 2017, a report by ProPublica exposed a controversial online function on Facebook that allowed advertisers to exclude members of its "ethnic affinity" groups, primarily people of color, from targeted marketing for certain ads.[8] Those ads were specifically focused on housing, employment, and the extension of credit.

In their controversies, Google explained their biases as problems associated with the algorithm or the inappropriate meta-tagging of images. Facebook immediately ended their practices and forbade advertisers from engaging in discriminatory practices on their site. In both cases, certain online users were wrongly characterized based upon their race.

We live in a society where online data are collected in real-time from users through a series of interactions with web sites, social media communities, e-commerce vehicles, and general online inquiries for information of interest. These small portions of data become compiled, mined, and eventually regenerated for commercial or public use. Big data serves a variety of purposes, from helping to advance breakthroughs in science, health care, energy, and transportation to enhancing government efficiencies by aggregating citizen input.

Big data can also exclude people. In a report published by the Federal Trade Commission (FTC), when big data analytics are misapplied, online users can be tracked or profiled based on their online

[6] BBC. (2013). Google searches expose racial bias, says study of names. *BBC*. Available at: https://goo.gl/P8oodF (last accessed June 25, 2019).
[7] Kasperkevic, J. (2015). Google says sorry for racist auto-tag in photo app. *The Guardian*. Available at: https://goo.gl/ZEJYng (last accessed June 25, 2019).
[8] Angwin, J. Tobin, A. Varner, M. (2017). Facebook (Still) Letting Housing Advertisers Exclude Users by Race. *ProPublica.* Available at: https://goo.gl/Vk4jrs (last accessed June 25, 2019).

activities and behaviors.[9] Consequently, online users can be denied credit based on their web browsing history, or aggregated, predictive analytics can wrongly determine an individual's suitability for future employment or an educational opportunity. Online proxies, including one's zip code, can also be used by marketers to extrapolate an individual's socioeconomic status based on neighborhood, resulting in incorrect assumptions about one's lifestyle or preferences.[10] In these and other examples, big data, when misapplied, can lead to the disparate treatment of individuals and groups, especially those that comprise protected classes by race, gender, age, ability, religion, and sexual orientation.

In these cases, the algorithm - when applied to these vulnerable populations - may repeat historical discrimination, or generate new forms of bias, whether explicit, implicit, or unconscious. In the instances of *explicit bias*, algorithms may not start out being discriminatory or have prejudicial intent. Instead, the algorithm can adapt to the societal biases that exist within communities of online users, leading to stereotypes and unfair profiling. Latanya Sweeney, Harvard researcher and former chief technology officer at the Federal Trade Commission (FTC), found the micro-targeting of higher-interest credit cards and other financial products when the computer inferred that the subjects were African-Americans, despite having similar backgrounds to whites.[11] During a public presentation at a FTC hearing on big data, Sweeney demonstrated how a web site, which marketed the centennial celebration of an all-black fraternity, received continuous ad suggestions for purchasing "arrest records" or accepting high-interest credit card offerings.[12]

---

[9] Ramirez, E. Brill, J. Ohlhausen, K. McSweeny, T. (2016). Big Data: A Tool for Inclusion or Exclusion. *FTC.* Available at: https://goo.gl/wUxwU1 (last accessed June 25, 2019).
[10] Noyes, K. (2015). Will big data help end discrimination—or make it worse? *Fortune.* Available at: https://goo.gl/VnPM1i (last accessed June 25, 2019).
[11] Sweeney, Latanya and Jinyan Zang. "How appropriate might big data analytics decisions be when placing ads?" Powerpoint presentation presented at the Big Data: A tool for inclusion or exclusion, Federal Trade Commission conference, Washington, DC. September 15, 2014. Available at https://www.ftc.gov/systems/files/documents/public_events/313371/bigdata-slides-sweeneyzang-9_15_14.pdf (last accessed June 25, 2019).
[12] "FTC Hearing #7: The Competition and Consumer Protection Issues of Algorithms, Artificial Intelligence, and Predictive Analytics," § Federal Trade Commission (2018),

When the values and beliefs of the programmer factors into the design of the algorithm, there is a risk of *implicit or unconscious biases*. Here, implicit bias can extend into the complex calculations of machine learning and artificial intelligence concealed within the design of the algorithmic procedure. Online retailer Amazon, whose global workforce is 60 percent male and where men hold 74 percent of the company's managerial positions, recently discontinued use of a recruiting algorithm after discovering gender bias.[13] The data that engineers used to create the algorithm were derived from the resumes submitted to Amazon over a 10-year period, which were predominantly from white males. The algorithm was taught to recognize word patterns in the resumes, rather than relevant skill sets, and these data were benchmarked against the company's predominantly male engineering department to determine an applicant's fit. As a result, the AI software penalized any resume that contained the word "women's" in the text and downgraded the resumes of women who attended women's colleges, resulting in gender bias.[14]

MIT researcher Joy Buolamwini found that the algorithms powering three commercially available facial recognition software systems were failing to recognize darker-skinned complexions.[15] Generally, most facial recognition training data sets are estimated to be more than 75 percent male and more than 80 percent white. When the person in the photo was a white man, the software was accurate 99 percent of the time at identifying the person as male. According to Buolamwini's research, the product error rates for the three products were less than one percent overall, but increased to more than 20

---

[13] Hamilton, Isobel Asher. "Why It's Totally Unsurprising That Amazon's Recruitment AI Was Biased against Women." Business Insider, October 13, 2018. Available at https://www.businessinsider.com/amazon-ai-biased-against-women-no-surprise-sandra-wachter-2018-10 (last accessed June 25, 2019).

[14] Vincent, James. "Amazon Reportedly Scraps Internal AI Recruiting Tool That Was Biased against Women." The Verge, October 10, 2018. Available at https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report (last accessed April 20, 2019). Although Amazon scrubbed the data of the particular references that appeared to discriminate against female candidates, there was no guarantee that the algorithm could not find other ways to sort and rank male candidates higher so it was scrapped by the company.

[15] Hardesty, Larry. "Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems." MIT News, February 11, 2018. Available at http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 (last accessed June 25, 2019). These companies were selected because they provided gender classification features in their software and the code was publicly available for testing.

percent in one product and 34 percent in the other two in the identification of darker-skinned women as female.[16] In response to Buolamwini's facial-analysis findings, both IBM and Microsoft committed to improving the accuracy of their recognition software for darker-skinned faces. Not surprising, Buolamwini is an African-American female researcher, suggesting that implicit and unconscious biases can often go undetected in high-tech industries where diverse populations are clearly underrepresented.

Generally, these examples of explicit, implicit and unconscious biases, complicated by historical realities, unmask the fact that algorithms are not necessarily devoid of societal biases, prejudices, stereotypes, and even incorrect assumptions.

Causes of algorithmic biases

Before delving into the specific use cases impacting the financial services sector, it is imperative to understand the root causes of online biases. *First*, historical human biases are shaped by pervasive and often deeply embedded prejudices against certain groups, which can lead to their reproduction and amplification in computer models. In the Amazon recruitment algorithm, men were the benchmark for professional "fit," resulting in female applicants and their attributes being downgraded. Unfortunately, historical realities often find their way into the algorithm's development and execution, and they are exacerbated by the lack of diversity which exists within the computer and data science fields.[17]

Second, online biases can also be reinforced and perpetuated without the user's knowledge. For example, African-Americans who are primarily the target for high-interest credit card options might find themselves clicking on this type of ad without realizing that they will continue to receive such predatory online suggestions. In this and other cases, the algorithm may never accumulate counter-factual ad

[16] Ibid.
[17] Turner Lee, Nicol. "Inclusion in Tech: How Diversity Benefits All Americans," § Subcommittee on Consumer Protection and Commerce, United States House Committee on Energy and Commerce (2019). Also available on Brookings web site, https://www.brookings.edu/testimonies/inclusion-in-tech-how-diversity-benefits-all-americans/ (last accessed June 25, 2019).

suggestions (e.g., lower-interest credit options) that the consumer could be eligible for and prefer. Thus, it is important for algorithm designers and operators to watch for such potential negative feedback loops that cause an algorithm to become increasingly biased over time.

Third, Insufficient training data is another cause of algorithmic bias, particularly if the data used to train the algorithm are more representative of some groups of people than others. In this case, the predictions from the model may be systematically worse for unrepresented or under-representative groups. For example, in Buolamwini's facial-analysis experiments, the poor recognition of darker-skinned faces was largely due to their statistical under-representation in the training data. Conversely, algorithms with too much data, or an over-representation, can skew the decision toward a particular result. Researchers at Georgetown Law School found that an estimated 117 million American adults are in facial recognition networks used by law enforcement, and that African-Americans were more likely to be singled out primarily because of their *over-representation* in mug-shot databases.[18] Consequently, African-American faces had more opportunities to be falsely matched, which produced a biased effect.

Managing bias detection and mitigating out biases

In our recent paper, the co-authors and I argue that detection approaches should begin with careful handling of the sensitive information of users, including data that identify a person's membership in a federally protected group (e.g., race, gender). Moreover, developers and other entities that are tasked in the design and deployment of algorithms must address and guard against the systemic bias waged on protected classes, especially when it leased to collective *disparate impacts*. Some of these outcomes may have a basis for legally cognizable harms, such as the denial of credit, online racial profiling, or mass surveillance.[19] While other cases may be justification for action simply due to the outputs of the

---

[18] Sydell, Laura. "It Ain't Me, Babe: Researchers Find Flaws In Police Facial Recognition Technology." NPR.org, October 25, 2016. Available at https://www.npr.org/sections/alltechconsidered/2016/10/25/499176469/it-aint-me-babe-researchers-find-flaws-in-police-facial-recognition (last accessed June 25, 2019).

[19] Guerin, Lisa. "Disparate Impact Discrimination." www.nolo.com. Available at https://www.nolo.com/legal-encyclopedia/disparate-impact-discrimination.htm (last accessed June 25, 2019). See also, Jewel v. NSA where the

algorithm, which may produce *unequal outcomes* or unequal error rates for different groups, despite not having an intent to discriminate, e.g., the mislabeling African-Americans as primates.

The argument could also be made that algorithms cannot be blind to sensitive attributes, despite all of the efforts of the developers. [20] Critics have pointed out that an algorithm may classify information based on online proxies for the sensitive attributes, yielding a bias against a group even without making decisions directly based on one's membership in that group. Barocas and Selbst define online proxies as "factors used in the scoring process of an algorithm which are mere stand-ins for protected groups, such as zip code as proxies for race, or height and weight as proxies for gender."[21] They argue that proxies often linked to algorithms can produce both errors and discriminatory outcomes, such as instances where a zip code is used to determine digital lending decisions or one's race triggers a disparate outcome.[22] Similarly, a job-matching algorithm may not receive the gender field as an input, but it may produce different match scores for two resumes that differ only in the substitution of the name "Mary" for "Mark" because the algorithm is trained to make these distinctions over time.

Going forward, operators of algorithms must be more transparent in their handling of sensitive information, especially if the potential proxy could itself be a legal classificatory harm.[23]

Addressing algorithmic bias in the financial services sector

For years, research has made clear that there are historical and contemporary inequalities that exist in the financial services industries. In the area of banking services for historically disadvantaged populations, the 2017 Federal Deposit Insurance Corporation's National Survey of Unbanked and

---

Electronic Frontier Foundation argues that massive (or dragnet) surveillance is illegal. Information about case available at https://www.eff.org/cases/jewel (last accessed June 25, 2019).

[20] This is often called an anti-classification criterion that the algorithm cannot classify based on membership in the protected or sensitive classes.

[21] Zarsky, Tal. "Understanding Discrimination in the Scored Society." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, January 15, 2015. https://papers.ssrn.com/abstract=2550248.

[22] Larson, Jeff, Surya Mattu, and Julia Angwin. "Unintended Consequences of Geographic Targeting." Technology Science, September 1, 2015. Available at https://techscience.org/a/2015090103/ (last accessed June 25, 2019).

[23] Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 2016. Available at https://papers.ssrn.com/abstract=2477899.

Underbanked Households reported that 17 percent of African-Americans and 14 percent of Hispanics were completely unbanked, compared to three percent of whites.[24] A further 30 percent of African-Americans and 29 percent of Hispanics were underbanked, compared to 14 percent of whites.

When coupled with other demographics, the disparities appear more glaring. Fifteen percent of unmarried female-headed family households are unbanked, as are 22 percent of American households without a high school diploma, and 24 percent of households where Spanish is the predominant language. As this evidence suggests, these populations are poorly served by the banking system as it currently operates.

African-Americans and non-White Hispanics are also poorly represented in homeownership. For example, Philadelphia has perhaps one of the most glaring displays of redlining, a practice which persuades and dissuades individuals toward change. Despite being part of 44 percent of the state's population, African-Americans received 10 times fewer mortgage loans than their white counterparts.

Despite a strengthening economy, record low unemployment and higher wages for whites, African-American homeownership has decreased every year since 2004 while all other groups have made gains. In 2017, 19.3 percent of African American applicants were denied home loans, while only 7.9 percent of white applicants were rejected. Brookings fellow Andre Perry found that "owner-occupied homes in black neighborhoods are undervalued by $48,000 per home on average, amounting to $156 billion in cumulative losses."[25] In other words, for every $100 in white family wealth, black families hold just $5.04. This type of physical redlining is now manifesting in the form of applications discrimination, or

---

[24] "FDIC National Survey of Unbanked and Underbanked Households." Federal Deposit Insurance Corporation. (2017). Available at economicinclusion.gov/downloads/2017_FDIC_Unbanked_HH_Survey_Appendix.pdf. (last accessed June 25, 2019).

[25] Perry, A., Rothwell, J., and Harshbarger, D. The devaluation of black assets in black neighborhoods: the case of residential property. Brookings. (2018). Available at https://www.brookings.edu/research/devaluation-of-assets-in-black-neighborhoods/ (last accessed June 25, 2019).

what Frank Pasquale has coined as “weblining,” where whole communities are classified by their credit characteristics and associated risks.

In his paper on credit denial in the age of AI, Brookings scholar Aaron Klein argues that AI and machine learning algorithms can begin to find “empirical relationships between new factors and consumer behaviors.”[26] In fact, he asserts that one’s social media profile, the type of computer one is using, what a person is wearing and where they buy they clothes could potentially factor into a credit model, denying loans to individuals whose choices and preferences suggest their inability to re-pay a loan. These new deployments of credit-algorithms are challenging legally cognizable harms and make it more difficult for consumers to discern the reasons for deniability.

Brookings scholar Henry-Nickie presents a similar argument in pointing to how the over-reliance on AI-driven financial services can create “wicked problems” when bank and fintech algorithms choose which consumers to serve.[27] In particular, the range of problems created by less thoughtful AI implementation can encompass: product steering, discriminatory pricing, unfair credit rationing, exclusionary filtering, and digital redlining.[28]

The historical and contemporary realities of certain populations when it comes to wealth- and asset-building suggest that more work needs to be done to avert a potential “double” and “triple” jeopardy of potential exclusion from the burgeoning online economy, particularly when algorithmic decision-making models are baked with assumptions about certain groups.

Recommendations

[26] Klein, A. Credit denial in the age of AI. Brookings. (2019). Available at https://www.brookings.edu/research/credit-denial-in-the-age-of-ai/ (last accessed June 25, 2019).
[27] Henry-Nickie, M. How artificial intelligence affects financial consumers. Brookings. (2019). Available at https://www.brookings.edu/research/how-artificial-intelligence-affects-financial-consumers/ (last accessed June 25, 2019).
[28] Ibid.

While this committee is embarking on both an educational pathway and serious legislative dialogue on the application of AI to financial services, I outline in my final section of this written testimony a set of high-level recommendations for consideration among Members of the committee and Congress as a whole.

1. **Congress must modernize civil rights laws and other consumer protections to safeguard protected classes from online discrimination.**

To develop trust from policymakers, computer programmers, businesses, and other operators of algorithms must abide by U.S. laws and statutes that currently forbid discrimination in public spaces. Historically, nondiscrimination laws and statutes unambiguously define the thresholds and parameters for the disparate treatment of protected classes. The 1964 Civil Rights Act "forbade discrimination on the basis of sex as well as race in hiring, promoting, and firing." The 1968 Fair Housing Act prohibits discrimination in the sale, rental, and financing of dwellings, and in other housing-related transactions to federally protected classes. Enacted in 1974, the Equal Credit Opportunity Act stops any creditor from discriminating against any applicant from any type of credit transaction based on protected characteristics. While these laws do not necessarily mitigate and resolve other implicit or unconscious biases that can be baked into algorithms, companies and other operators should guard against violating these statutory guardrails in the design of algorithms, as well as mitigating their implicit concern to prevent past discrimination from continuing.

To quell algorithmic bias, Congress should start by clarifying how these nondiscrimination laws apply to the types of grievances recently found in the digital space, since most of these laws were written before the advent of the internet.[29] Such legislative action can provide clearer guardrails that are triggered when algorithms are contributing to legally recognizable harms. Moreover, when creators and

[29] Tobin, Ariana. "HUD sues Facebook over housing discrimination and says the company's algorithms have made the problem worse." ProPublica (March 28, 2019). Available at https://www.propublica.org/article/hud-sues-facebook-housing-discrimination-advertising-algorithms (last accessed June 25, 2019).

operators of algorithms understand that these may be more or less non-negotiable factors, the technical design will be more thoughtful in moving away from models that may trigger and exacerbate explicit discrimination, such as design frames that exclude rather than include certain inputs or are not checked for bias.[30]

Henry-Nickie also points to the importance of maintaining consumer financial protections in the age of AI, which implore regulators to engage in targeted, strategic and analytical exploration of emerging technologies in the sector. Further, she argues that "the deluge of data generated by connected devices and machine learning applications creates a prime opportunity to collect and mine publicly available data to inform critical regulation burden analyses," for which she references the Small Business Regulatory Enforcement Fairness Act and the Paperwork Reduction Act.[31]

In the end, it is important for Congress to determine what role, if any, they want to play in prescribing some level of accountability to companies developing and disseminating algorithms going forward. It may be the case that without accountability or further conversation between policymakers, technologists and civil society, this conversation will be for naught.

2. **Companies that design and deploy algorithms must exercise some level of algorithmic accountability, which involves the creation of a bias impact statement, regular auditing and more human involvement in risk-adverse decisions, like credit and lending.**

As a self-regulatory practice, a *bias impact statement* can help probe and avert any potential biases that are baked into or are resultant from the algorithmic decision. As a best practice, operators of algorithms should brainstorm a core set of initial assumptions about the algorithm's purpose prior to its development and execution. The bias impact statement should assess the algorithm's purpose, process

---

[30] Elejalde-Ruiz, Alexia. "The end of the resume? Hiring is in the midst of technological revolution with algorithms, chatbots." Chicago Tribune (July 19, 2018). Available at http://www.chicagotribune.com/business/ct-biz-artificial-intelligence-hiring-20180719-story.html.

[31] Henry-Nickie, M. How artificial intelligence affects financial consumers. Brookings. (2019). Available at https://www.brookings.edu/research/how-artificial-intelligence-affects-financial-consumers/ (last accessed June 25, 2019).

and production, where appropriate. Operators of algorithms should also consider the role of diversity within their work teams, training data, and the level of cultural sensitivity within their decision-making processes. Employing diversity in the design of algorithms upfront will trigger and potentially avoid harmful discriminatory effects on certain protected groups, especially racial and ethnic minorities. While the immediate consequences of biases in these areas may be small, the sheer quantity of digital interactions and inferences can amount to a new form of systemic bias. Therefore, the operators of algorithms should not discount the possibility or prevalence of bias and should seek to have a diverse workforce developing the algorithm, integrate inclusive spaces within their products, or employ "diversity-in-design," where deliberate and transparent actions will be taken to ensure that cultural biases and stereotypes are addressed upfront and appropriately. Adding inclusivity into the algorithm's design can potentially vet the cultural inclusivity and sensitivity of the algorithms for various groups and help companies avoid what can be litigious and embarrassing algorithmic outcomes.

The bias impact statement should not be an exhaustive tool. As a self-regulatory tool, its goal should be to ward off disparate impacts resulting from the algorithm that border on unethical, unfair, and unjust decision-making. When the process of identifying and forecasting the purpose of the algorithm is achieved, a robust feedback loop will aid in the detection of bias, which leads to the next recommendation of promoting regular audits of algorithms and their decisions. Where appropriate, more humans should be involved in these processes to ensure that subjective criteria are not dominating the final outcome.

Congress should promote, and in some cases reward self-regulatory models where businesses identify, monitor and correct biases that negatively impact the online experiences of users. For example, Google's decision to ban ads that promoted payday loans was an example of self-regulation. Or, Facebook's updates to its ad policies to prevent race-based targeting, especially those that attempt to include or exclude demographic groups in housing, employment and credit, is another example of how companies are correcting ill-advised practices. A potentially novel idea may be to reward best practices with some type of "gold seal of approval" when companies demonstrate a strict adherence to standards and practices which highlights their outperformance in creating more ethical algorithms.

3. **Congress should support the use of regulatory sandboxes and safe harbors to curb online biases.**

Regulatory sandboxes could be another policy strategy for the creation of temporary reprieves from regulation to allow the technology and rules surrounding its use to evolve together. These policies could apply to algorithmic bias and other areas where the technology in question has no analog covered by existing regulations. Rather than broaden the scope of existing regulations or create rules in anticipation of potential harms, a sandbox allows for innovation both in technology and its regulation. Even in a highly regulated industry, the creation of sandboxes where innovations can be tested alongside with lighter touch regulations can yield benefits.

For example, companies within the financial sector that are leveraging technology, or fintech, have shown how regulatory sandboxes can spur innovation in the development of new products and services.[32] These companies make extensive use of algorithms for everything from spotting fraud to deciding to extend credit. Some of these activities mirror those of regular banks, and those would still fall under existing rules, but new ways of approaching tasks would be allowed within the sandbox.[33] Because sandboxes give innovators greater leeway in developing new products and services, they will require active oversight until technology and regulations mature. The U.S. Treasury recently reported not only on the benefits that countries that have adopted fintech regulatory sandboxes have realized,

---

[32] Fintech regulatory sandboxes in UK, Singapore, and states in the U.S. are beginning to authorize them. They allow freedom to offer new financial products and use new technologies such as blockchain.

[33] In March, the state of Arizona became the first U.S. state to create a "regulatory sandbox" for fintech companies, allowing them to test financial products on customers with lighter regulations. The U.K. has run a similar initiative called Project Innovate since 2014. The application of a sandbox can allow both startup companies and incumbent banks to experiment with more innovative products without worrying about how to reconcile them with existing rules.

but recommended that the U.S. adopt fintech sandboxes to spur innovation.[34] Given the broad usefulness of algorithms to spur innovation in various regulated industries, participants in the roundtables considered the potential usefulness of extending regulatory sandboxes to other areas where algorithms can help to spur innovations.

Regulatory safe harbors could also be employed, where a regulator could specify which activities do not violate existing regulations.[35] This approach has the advantage of increasing regulatory certainty for algorithm developers and operators. For example, Section 230 of the Communications Decency Act removed liability from websites for the actions of their users, a provision widely credited with the growth of internet companies like Facebook and Google. The exemption later narrowed to exclude sex trafficking with the passage of the Stop Enabling Online Sex Trafficking Act and Fight Online Sex Trafficking Act. Applying a similar approach to algorithms could exempt their operators from liabilities in certain contexts while still upholding protections in others where harms are easier to identify. In line with the previous discussion on the use of certain protected attributes, safe harbors could be considered in instances where the collection of sensitive personal information is used for the specific purposes of anti-bias detection and mitigation.

4. **The tech sector must be more deliberate and systematic in the recruitment, hiring and retention of diverse talent to avert and address the mishaps generated by online discrimination, especially algorithmic bias.**

Less diverse workforces contribute to algorithmic bias, whether intentional or not. Recent diversity statistics report these companies employ less than two percent of African Americans in senior executive

[34] Mnuchin, Steven T., and Craig S. Phillips. "A Financial System That Creates Economic Opportunities - Nonbank Financials, Fintech, and Innovation." Washington, D.C.: U.S. Department of the Treasury, July 2018.Available at https://home.treasury.gov/sites/default/files/2018-08/A-Financial-System-that-Creates-Economic-Opportunities---Nonbank-Financials-Fintech-and-Innovation_0.pdf (last accessed June 25, 2019).

[35] Another major tech-related Safe Harbor is the EU-US Privacy Shield after the previous Safe Harbor was declared invalid in the EU. Available at https://en.wikipedia.org/wiki/EU%E2%80%93US_Privacy_Shield (last accessed June 25, 2019).

positions, and three percent of Hispanics when compared to 83 percent of whites.[36] Asian-Americans comprise just 11 percent of executives in high tech companies.[37] In the occupations of computer programmers, software developers, database administrators, and even data scientists, African-Americans and Hispanics collectively are under six percent of the total workforce, while whites make up 68 percent.[38] Even when people of color are employed in high tech industries, the feelings of professional and social isolation also have been shown to marginalize these employees, potentially restricting their active workplace engagement, affecting their participation in the feedback loop, and contributing to higher rates of attrition.[39] At Google, employees have been subjected to anti-diversity memos,[40] and women have experienced documented backlash from male employees on hiring. This alienation within high-tech workforces neither encourages nor welcomes diverse input into work products. It also may distract from efforts to incorporate elements of "diversity in the design" of algorithms, where biases can be avoided at the onset. Technologists may not be necessarily trained to identify cues that are outside of their cultural context and can be fenced into work groups that share similar experiences, values and beliefs. This is what some researchers have dubbed *inattentional blindness*.

These largely unconscious bias errors strongly support why high-tech companies should be striving for more diverse workforces to identify and quell online discrimination. Companies that are disrupting societal norms through the sharing economy, social media and the internet of things must do better to

---

[36] Atwell, J. (2016). Lack of women and minorities in senior investment roles at venture capital firms. *Deloitte.* Available at: https://goo.gl/iah1VZ (accessed June 25, 2019).
[37] Ibid.
[38] EEOC. (2016). Diversity in High Tech. *EEOC*. Available at: https://goo.gl/EwKBUJ (accessed June 25, 2019).
[39] Scott, A. Kapor Klein, F. Onovakpuri,U. (2017). Tech Leavers Study. *Kapor Center*. Available at: https://goo.gl/Zgf6dg (accessed June 25, 2019).
[40] Conger, K. (2017). Here's the 10-page anti-diversity screed circulating internally at Google. *Gizmodo*. https://goo.gl/UEYNhx. Available at: https://goo.gl/9ctiyF (accessed June 25, 2019).

address the less than remarkable representation of people of color as creators, influencers and decision makers.

As in the case of HBCUs and HSIs, the tech sector should work to strengthen those relationships and programs, which target these students for future employment. Congress and federal agencies, including the U.S. Department of Education, need to also do more to ensure that minority-serving institutions are establishing premiere programs that include both technology access and cutting-edge career development in fields where the nation will soon face massive shortages. We need to take notes from the former Obama administration that pushed the U.S. toward a “race to the top,” urging collaboration between the private and public sectors to realize the nation’s global competitiveness and edge over our international counterparts.

Conclusion

The prevalence of AI and machine learning should trigger alarms when we fail to have collaborative, proactive and productive discussions on their applications design and use. While many innocuous decisions will be best served by algorithms, others, especially those emanating from the financial services industry, may need more thoughtful consideration on their intended and unintended consequences. For developers seeking to deploy these emerging technologies, the engagement in conversations about its responsible and ethnical deployment are at the core of these conversations, and potentially result in reduced risk to consumers and more deliberation in the identification and mitigation of online biases.

I want to thank Members of this Committee for including me in this conversation and look forward to your questions.