GEORGETOWN LAW
Center on Privacy & Technology

September 10, 2021

National Institute of Standards and Technology
Attn: Information Technology Laboratory
100 Bureau Drive
Gaithersburg, MD 20899-2000

VIA EMAIL
ai-bias@list.nist.gov

Re: Comments of Georgetown Center on Privacy & Technology in response to Draft NIST Special Publication 1270: A Proposal for Identifying and Managing Bias in Artificial Intelligence (June 2021)

The Center on Privacy & Technology at Georgetown Law (the "Center") is pleased to submit comments on the National Institute of Standards and Technology's (NIST) Draft NIST Special Publication 1270: A Proposal for Identifying and Managing Bias in Artificial Intelligence (the "draft proposal"). It is encouraging that NIST is engaging with the issue of algorithmic bias, and to see the draft proposal acknowledge the importance of pre-design considerations and the role of societal context surrounding algorithms. That said, this submission will present some key concerns and propose four recommendations to address them. We hope our comments might also inform NIST's future work on algorithmic bias and bias in artificial intelligence (AI).[1]

Our four recommendations are:

1) **Ensure that future and final proposals and standards center civil rights impacts and retain explicit consideration of the sociotechnological, historical, and political contexts in which algorithms are embedded.** Although the draft proposal acknowledges issues of societal context and power related to algorithmic bias, the conclusion ultimately reverts to a potentially technologically deterministic model by focusing on technical requirements to address such bias. Truly resolving the issues at the heart of algorithmic discrimination requires understanding and addressing such discrimination as inseparable from the broader sociopolitical, historical, and legal contexts that shape and are shaped by the development of AI tools and systems. These contexts include the perspectives and values of those who create, sell, and use such systems, relative to those of the communities who are made the subjects of algorithmically biased systems.

[1] The terms "algorithmic bias", "AI bias", and "bias in AI" will be used interchangeably throughout this submission.

2) **Emphasize pre-design considerations early on, such as the purpose of a proposed AI tool or system and avoiding technological solutionism, over more subjective factors such as "public trust."** Undue focus on public trust as a measure of acceptability substitutes public opinion for rigorous, evidence-based, and civil rights-centering oversight mechanisms and evaluation of an algorithm's impact.

3) **Incorporate analysis and recommendations from privacy law scholarship regarding the dangers of "managerialization" and legal endogeneity of privacy compliance, and apply them to the algorithmic bias context.** An approach focused on "managing" bias may result in a legal regime that perpetuates rather than mitigates or eliminates algorithmic harms to historically marginalized and vulnerable communities.

4) **Add a fourth stage to the "AI lifecycle": post-deployment, which should have its own set of obligations distinct from those in the active deployment stage and its immediate aftermath.** Post-deployment refers to the stage after the algorithmic tool or system has already been implemented, when it is simply "running." This stage should include certain obligations for as long as the AI tool or system remains in place, such as periodic testing and audits, updated algorithmic impact assessments, complaint and recall mechanisms, and regular public justification for continued use.

The remainder of this submission will further discuss each of the above recommendations in turn.

Recommendation 1: Ensure that future and final proposals and standards center civil rights impacts and retain explicit consideration of the sociotechnological, historical, and political contexts in which algorithms are embedded.

We were encouraged to see the draft proposal recognize that algorithmic bias issues implicate broader societal and contextual considerations that go beyond the technical. We emphasize that NIST should ensure all future and final versions of this proposal, and any eventual standards, expressly integrate a sociotechnological lens and center civil rights impacts throughout all stages of the "AI lifecycle." Further, such proposals and standards should acknowledge explicitly that although technical standards may be necessary, they are seldom sufficient to prevent, detect, or address algorithmic bias. To support this overarching priority recommendation, we provide three supplementary recommendations:

1a) Integrate Frameworks from "Critical AI" Scholarship

We strongly urge NIST to consult and cite further scholarship that confronts algorithmic bias as an issue of sociotechnical systems, systemic oppression, historical and political context, and power — what might be thought of as "critical AI" scholarship. Such scholarship moves beyond the conventional AI "fairness, accountability, transparency, (ethics)" frame to tackle more difficult and searching questions that implicate the field of algorithmic accountability itself, and which are necessary to raise in order to get to the heart of the problem. The following papers, for example, all discuss this aspect of the algorithmic bias problem in greater depth and set out concrete recommendations and approaches that NIST might consider adopting (alphabetical by title):[2]

- "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence," by Shakir Mohamed, Marie-Therese Png, and William Isaac;[3]
- "Studying up: reorienting the study of algorithmic fairness around issues of power," by Chelsea Barabas, Colin Doyle, J.B. Rubinovitz, and Karthik Dinakar;[4] and
- "Towards a Critical Race Methodology in Algorithmic Fairness," by Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud.[5]

1b) Build Sociotechnological and Civil Rights Expertise within Federal Agencies

We also urge NIST and the Department of Commerce more broadly to support efforts to build internal sociotechnological and civil rights expertise, such as creating offices or internal team units dedicated to technology and civil rights and staffing them with legal and other experts, advocates, and activists who have been immersed in the intersection of technology and civil rights issues. This includes supporting efforts to build such teams within agencies that are in a position to oversee the use of algorithmic decision-making systems in specific sectors. Laura Moy (the Center's Associate Director) and Gabrielle Rejouis have set out in a joint paper specific recommendations for how to build such capacity within the federal government.[6]

Algorithmic bias exacerbates discrimination in areas such as housing, employment, and financial services, for instance[7] — areas which there are already one or more dedicated federal agencies responsible for regulating. However, these agencies to date have not met or been equipped to meet the challenges that algorithmic bias poses in their respective domains, despite considerable harms that have occurred and continue to occur. For example:

- Algorithmic accountability researchers and workers' rights advocates have shown that the Department of Labor, Equal Employment Opportunity Commission (EEOC), Office of Federal Contract Compliance Programs, and related agencies must do more to

---

2 These are listed in addition to a paper which is already referenced in the draft proposal: Andrew D. Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, & Janet Vertesi, *Fairness and Abstraction in Sociotechnical Systems*, delivered at *Conference on Fairness, Accountability, and Transparency (FAT* '19)*, January 29-31, 2019, Atlanta, GA, http://sorelle.friedler.net/papers/sts_fat2019.pdf.

3 Shakir Mohamed, Marie-Therese Png & William Isaac, *Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence*, 33 Philosophy & Technology 659 (2020), https://arxiv.org/pdf/2007.04068.pdf.

4 Chelsea Barabas, Colin Doyle, J.B. Rubinovitz & Karthik Dinakar, *Studying up: reorienting the study of algorithmic fairness around issues of power*, delivered at *Conference on Fairness, Accountability, and Transparency (FAT* '20)*, January 27-30, 2020, Barcelona, Spain, https://dl.acm.org/doi/pdf/10.1145/3351095.3372859.

5 Alex Hanna, Emily Denton, Andrew Smart & Jamila Smith-Loud, *Towards a Critical Race Methodology in Algorithmic Fairness*, delivered at *Conference on Fairness, Accountability, and Transparency (FAT* '20)*, January 27-30, 2020, Barcelona, Spain, https://arxiv.org/pdf/1912.03593.pdf.

6 Laura Moy & Gabrielle Rejouis, *Addressing Challenges at the Intersection of Civil Rights and Technology*, (December 2020), https://9381c384-0c59-41d7-bbdf-62bbf54449a6.filesusr.com/ugd/14d834_8919b8318a674ff79dcf8d7433907c8b.pdf.

7 "Letter to White House OSTP on Centering Civil Rights in AI Policy," ACLU (July 13, 2021), https://www.aclu.org/sites/default/files/field_document/2021-07-13_letter_to_white_house_ostp_on_centering_civil_rights_in_ai_policy_1.pdf.

effectively address algorithmic bias in hiring and other workplace technology issues,[8] including algorithmic management.[9] Experts have recommended, for instance, that the EEOC proactively investigate algorithmic hiring tools and issue guidelines that explicitly interpret non-discrimination laws to apply to employers' use of algorithmic decision-making tools.[10]

- The Department of Housing and Urban Development (HUD) faced strong criticism of its final rule on the Fair Housing Act's disparate impact standard because the rule "makes it nearly impossible for victims of algorithmic discrimination to hold companies accountable, and encourages housing providers to adopt and use discriminatory algorithms."[11] The final rule reflects lack of understanding of the role of algorithmic tools in housing decisions, specifically how they intersect with historical and systemic racial discrimination in housing, and the high evidentiary threshold required to support a claim fails to take into account the position of impacted communities on the ground.[12]

  Moreover, the HUD had already dispensed with an even more controversial and harmful rule proposed earlier, only after hearing from "a coalition of 23 civil rights and consumer advocacy organizations and individual experts [and] 45,000 comments from civil and human rights organizations, data scientists, housing and financial services providers, disability rights groups, and more."[13] Such expertise should already be internal to agencies such as HUD as a matter of course, in order to fulfill their mandates as applied to 21st-century situations, rather than relying on often under-resourced non-

---

8 See e.g., Miranda Bogen & Aaron Rieke, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, Upturn (December 2018), https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf; Aaron Rieke, Urmila Janardan, Mingwei Hsu & Natasha Duarte, *Essential Work: Analyzing the Hiring Technologies of Large Hourly Employers*, Upturn (May 2021), https://www.upturn.org/reports/2021/essential-work/; and Lydia X. Z. Brown, Ridhi Shetty & Michelle Richardson, *Report — Algorithm-driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination?*, Center for Democracy & Technology (December 2020), https://cdt.org/insights/report-algorithm-driven-hiring-tools-innovative-recruitment-or-expedited-disability-discrimination/.

9 See e.g., *Put Workers over Profits: End Worker Surveillance*, Athena (October 14, 2020), https://athenaforall.medium.com/end-worker-surveillance-d99aa7cd3850; and Aiha Nguyen, *The Constant Boss: Work Under Digital Surveillance*, Data & Society (May 2021), https://datasociety.net/wp-content/uploads/2021/05/The_Constant_Boss.pdf.

10 Miranda Bogen & Aaron Rieke, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, Upturn (December 2018) at 46, https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf; Aaron Rieke, Urmila Janardan, Mingwei Hsu & Natasha Duarte, *Essential Work: Analyzing the Hiring Technologies of Large Hourly Employers*, Upturn (May 2021) at 41-42, https://www.upturn.org/reports/2021/essential-work/; and Lydia X. Z. Brown, Ridhi Shetty & Michelle Richardson, *Report — Algorithm-driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination?*, Center for Democracy & Technology (December 2020) at 19, https://cdt.org/insights/report-algorithm-driven-hiring-tools-innovative-recruitment-or-expedited-disability-discrimination/.

11 Lauren Sarkesian & Spandana Singh, *HUD's New Rule Paves the Way for Rampant Algorithmic Discrimination in Housing Decisions*, Open Technology Institute (October 1, 2020), https://www.newamerica.org/oti/blog/huds-new-rule-paves-the-way-for-rampant-algorithmic-discrimination-in-housing-decisions/

12 *Ibid*.

13 *Ibid*.

profit organizations, academics, and civil rights advocates to supply missing critical perspectives and expertise.

- Technology and civil rights groups have called for the Federal Trade Commission (FTC) to create an Office of Civil Rights to better understand and enforce against instances or patterns of algorithmic bias that constitute unfair and deceptive commercial data practices.[14] They have also urged the White House Office of Science and Technology Policy to center civil rights in AI and technology policy, and made agency-specific recommendations concerning algorithmic discrimination in housing, hiring, and financial services.[15]

Without these agencies and their counterparts in other sectors actively enforcing NIST's or other eventual standards or approaches to algorithmic bias, such an approach or set of standards may have limited impact on the ground. The Center thus urges NIST to build further expertise and capacity at the intersection of technology, algorithmic accountability, and civil rights within itself, and to support efforts to do the same in other federal agencies across the board.

1c) Understand the Risks of a Technical Solutions Framework to Address Algorithmic Bias

As mentioned, it is encouraging that the draft proposal recognizes the importance of "consider[ing] AI within the social system it operates" in (page 5)[16] and acknowledges the dangers of approaches that amount to technological solutionism or fail to take into account the role of power and social inequity. However, we are concerned that the proposal, towards the end or in future iterations, reverts or may revert to framing that centers technical requirements as a meaningful way to address algorithmic bias.

If the proposal or resulting standards could be read to endorse predominantly technical solutions, that may have the perverse effect of doing practical harm, rather than good, by enabling vendors, developers, and users of algorithmic decision-making tools to remain relatively ignorant of the contextual landscapes in which these tools will be used.[17] Too heavy a focus on technical solutions could cause such entities and regulators to miss fundamental issues such as who built or deployed an AI tool or system to begin with, for what purpose, based on what implicit values, with what long-term impacts on which communities, and to whose benefit (in practice, as opposed to in advertised intent).[18]

---

[14] Ian Weiner, *Federal Trade Commission Must Protect Civil Rights, Privacy in Online Commerce,* Lawyers' Committee for Civil Rights Under Law (August 4, 2021), https://www.lawyerscommittee.org/federal-trade-commission-must-protect-civil-rights-privacy-in-online-commerce/.

[15] "Letter to White House OSTP on Centering Civil Rights in AI Policy," ACLU (July 13, 2021), https://www.aclu.org/sites/default/files/field_document/2021-07-13_letter_to_white_house_ostp_on_centering_civil_rights_in_ai_policy_1.pdf.

[16] Page numbers referenced in-text or otherwise unattributed to a specific source all refer to the draft proposal.

[17] "[N]arrow technical conceptualizations of algorithmic fairness elide more fundamental issues and, in the process, run the risk of legitimizing harmful practices based on fundamentally unsound truth claims about the world." Chelsea Barabas, Colin Doyle, J.B. Rubinovitz & Karthik Dinakar, *Studying up: reorienting the study of algorithmic fairness around issues of power*, delivered at *Conference on Fairness, Accountability, and Transparency (FAT* '20)*, January 27-30, 2020, Barcelona, Spain, https://dl.acm.org/doi/pdf/10.1145/3351095.3372859 at 167 (inline citations omitted).

[18] See e.g., the following open letter and its extensive footnotes for a thorough explanation and example of this phenomenon in the context of algorithms that purport to predict "criminality" based on biometrics

NIST should ensure that its final version of the proposal, and any eventual standards or "risk management" approach expressly reflect the fact that "[f]airness and justice are properties of social and legal systems like employment and criminal justice, *not properties of the technical tools within*."[19] It is not possible to technically standardize or code one's way to the eradication of racism or income inequality.

To further emphasize this point, we present three example limitations of a technical approach to algorithmic bias:

i. Lack of Data versus Biased Data

Technical standards cannot, or should not, address algorithmic bias where the specific bias involved arises primarily from *lack of* data, rather than a skewed algorithm or discriminatory datasets (though all three may occur simultaneously in a given case).[20]

For example, a May 2021 large-scale study demonstrated that when algorithms discriminate against racialized and low-income applicants in determining mortgage approvals, that disadvantage is due to these groups of applicants having on average less credit history data than advantaged groups. This resulted in less precise predictions, which meant a relatively higher rate of false rejections.[21]

The "missing" data simply does not exist, due to redlining and other historical and present-day forms of discrimination preventing members of racialized and low-income communities from engaging in the activities that would have generated that data. Implementing technical standards for developing mortgage approval algorithms would do nothing to remove the financial and credit barriers, rooted in historical and present-day racist and classist policies, that ultimately lead to such discriminatory results.

Lack of data in an algorithmic decision-making system negatively impacting historically marginalized groups does not mean that the solution to algorithmic bias is to collect more data about those groups. On the contrary, in a world where Black and Indigenous peoples, immigrants, 2SLGBTQIA+ individuals, people with disabilities, and those with low-income or who are unhoused are already disproportionately and wrongly subjected to punitive surveillance, calls for even greater levels of data collection about them can amount to a form of "predatory inclusion." Predatory inclusion refers to the practice of including historically

---

and other forms of personal data: *Abolish the #TechToPrisonPipeline*, Coalition for Critical Technology (June 23, 2020), https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16.

19 Andrew D. Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, & Janet Vertesi, *Fairness and Abstraction in Sociotechnical Systems*, delivered at *Conference on Fairness, Accountability, and Transparency (FAT* '19)*, January 29-31, 2019, Atlanta, GA, http://sorelle.friedler.net/papers/sts_fat2019.pdf at 1 (emphasis added).

20 This issue might be considered a form of sampling or representation bias as defined in Appendix A, Table 1 of the draft proposal. However, in the case of this example, a "better sample" does not exist, meaning the bias cannot be removed, and even where more representative samples are possible through more expansive collection of data, that is an inadvisable response due to likely constituting a form of predatory inclusion, as explained later in this section.

21 Will Douglas Heaven, "Bias isn't the only problem with credit scores — and no, AI can't help," *MIT Technology Review* (June 17, 2021), https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/.

marginalized communities in a system, ostensibly to advance equality and democracy, but in a way that ultimately does such groups collective and long-term harm.[22]

ii. Questionable Purpose Regardless of Bias

Even the most rigorous technical standards mean little for upholding civil rights or the public interest where an algorithmic program serves a questionable purpose in the first place.

For example, the burgeoning field of "legal tech" includes a prominent area of work on algorithms that purport to predict how a certain judge or particular court may decide a given case.[23] One can imagine various kinds of biases that may appear in such algorithms, reflecting historical and present-day discrimination and unconscious biases embedded in judges, courts, precedents, and the law itself. Frank Pasquale, a law professor and author of *The Black Box Society,* notes, "Even in these relatively sedate areas of practice ["tax, will preparation, and traffic disputes"], [legal algorithms] have raised serious ethical concerns about unintended consequences and consumer protection."[24]

However, even if it were possible to technologically remove all biases from legal algorithms, technical standards cannot answer the much thornier philosophical and sociolegal question of whether any part of the legal system *should* rely on predictive algorithms rather than human judges.[25] Such a decision should emerge from society-wide democratic dialogue that occurs at the start of or even prior to the "AI lifecycle" of any such proposed algorithmic legal tools, and extends far beyond the purview of technical standards.

---

22 "Predatory inclusion is the logic, organization, and technique of including marginalized consumer-citizens into ostensibly democratizing mobility schemes on extractive terms." Tressie Mcmillan Cottom, *Where Platform Capitalism and Racial Capitalism Meet: The Sociology of Race and Racism in the Digital Society*, 6:4 Sociology of Race and Ethnicity 441 (2020) at 443.

23 See e.g., Matthew Hutson, "Artificial intelligence prevails at predicting Supreme Court decisions," *Science* (May 2, 2017), https://www.science.org/news/2017/05/artificial-intelligence-prevails-predicting-supreme-court-decisions; and Roy Strom, "Keep Judges and Lawyers Out of Legal Predictions, Tech CEO Says," *Bloomberg Law* (September 5, 2019), https://news.bloomberglaw.com/us-law-week/keep-judges-and-lawyers-out-of-legal-predictions-tech-ceo-says.

24 Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87:1 George Washington Law Review (2019) at 18, https://www.gwlr.org/wp-content/uploads/2019/01/87-Geo.-Wash.-L.-Rev.-1.pdf.

25 Legal scholars Ian Kerr and Carissima Mathen engaged in a thought experiment militating towards answering no: "[W]e have argued that legal reasoning cannot be reduced to mere functional capabilities regarding extraordinary information gathering, speed, memory, recall and even the ability to distinguish and disambiguate relevant legal rules. Legal reasoning, indeed, being a judge, requires the ability to *meaningfully follow* rules and to adopt a particular point of view of a legal system. Legal reasoning also requires being a *member of the community*, understanding its history, its moral convictions, having a point of view about its current character and having a stake in its future. On these foundational abilities, we have tried to articulate why [an exact robot replica of Supreme Court of the United States Chief Justice John Roberts] most likely did not qualify." Ian Kerr & Carissima Mathen, *Chief Justice John Roberts Is a Robot*, delivered at *We Robot*, April 4-5, Coral Gables, FL, at 39-40 (emphasis in original), http://robots.law.miami.edu/2014/wp-content/uploads/2013/06/Chief-Justice-John-Roberts-is-a-Robot-March-13-.pdf.

iii. Bias Entrenched by Algorithm's Very Existence

Technical standards are insufficient to address algorithmic bias issues when discriminatory harms result from the *existence and use of the algorithm at all*, rather than necessarily or exclusively bias within any part of the algorithm itself or the data on which it was trained.

The foremost example of this dynamic is so-called predictive policing technologies and other forms of algorithmic policing tools, including facial recognition software. Bias and discrimination embedded in these technologies, against historically marginalized groups, has been well documented.[26]

No matter how rigorously technical standards are enforced with such tools and programs, and even if they somehow become perfectly unbiased, algorithmic policing technologies are inseparable from their surrounding context.[27] This context comprises racist and discriminatory policing; state-sanctioned violence against Black and Indigenous peoples; the carceral legal system as upheld through the criminal law and by law enforcement; the criminalization of poverty; and systemic racism and other forms of systemic oppression perpetuated throughout every level of the criminal justice system.

Within such a context, the decision to use a hypothetical unbiased algorithm targeting, for instance, street-level car thefts, by its very existence, subjects offenders — or perceived potential offenders (with all the discriminatory implications) — of such thefts to disproportionately greater police scrutiny than offenders of, for instance, environmental crime or corporate negligence.[28] After all, no predictive policing tool is being equally used to target perpetrators of

---

26 See e.g., Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 Proceedings of Machine Learning Research 1 (2018); Kristian Lum & William Isaac, To predict and serve? 13:5 Significance 14 (2016); Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, "Machine Bias," *ProPublica* (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Andrew Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement* (2017); *Dismantling Predictive Policing in Los Angeles*, Stop LAPD Spying Coalition (May 2018), https://stoplapdspying.org/wp-content/uploads/2018/05/Before-the-Bullet-Hits-the-Body-May-8-2018.pdf; David Robinson & Logan Koepke, *Stuck in a Pattern*, Upturn (August 2016), https://www.upturn.org/reports/2016/stuck-in-a-pattern/; and Clare Garvie, Alvaro Bedoya & Jonathan Frankle, *The Perpetual Line-Up: Unregulated Police Face Recognition in America*, Center on Privacy & Technology at Georgetown Law (October 2016), https://www.perpetuallineup.org.

27 "Machine learning programs are not neutral; research agendas and the data sets they work with often inherit dominant cultural beliefs about the world. These research agendas reflect the incentives and perspectives of those in the privileged position of developing machine learning models, and the data on which they rely. The uncritical acceptance of default assumptions inevitably leads to discriminatory design in algorithmic systems, reproducing ideas which normalize social hierarchies and legitimize violence against marginalized groups." *Abolish the #TechToPrisonPipeline*, Coalition for Critical Technology (June 23, 2020), https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16 (citations omitted).

28 Kate Robertson, Cynthia Khoo & Yolanda Song, *To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada*, The Citizen Lab (2020) at 115, https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf. See also: "A more equitable application of the [LAPD's "Los Angeles Strategic Extraction and Restoration" (LASER)] program, if it were even possible, should not be the goal then. Such colorblind policies are not sufficient to adequately address institutionalized racism, which the LASER program is but one manifestation of." *Before the Bullet Hits the Body: Dismantling Predictive Policing in Los Angeles*, Stop LAPD Spying Coalition (May 2018), https://stoplapdspying.org/wp-content/uploads/2018/05/Before-the-Bullet-Hits-the-Body-May-8-2018.pdf.

or justify police intervention in the lives of potential perpetrators of the latter crimes.[29]

Recommendation 2: Emphasize pre-design considerations early on, such as the purpose of a proposed AI tool or system and avoiding technological solutionism, over more subjective factors such as "public trust."

Evidence has shown that AI tools and algorithmic decision-making systems that are "fraudulent, pseudoscientific, prey on the user, or generally exaggerate claims" may be "extreme" (page 7) in their harmful consequences, but are unfortunately common cases otherwise.[30] The draft proposal should reflect this state of affairs, and avoid language that suggests outright rejection of projects may be an exception rather than the rule. The list of tools and apps that warrant outright rejection should also include AI systems that *enable their users to prey on other people (including non-users)*, such as a deepfake algorithm on Telegram that enabled users to create fake nude photos of women based on single clothed photos of them.[31]

The draft proposal takes "engender[ing] public trust" in (page 4) and "improving acceptance of" (page 5) AI systems for granted as a desirable goal in and of itself, with seemingly disproportionate focus on "public trust" and acceptance. Putting such weight on the question of whether the public "trusts" or "accepts" AI distracts from the responsibility to evaluate a given AI system objectively in terms of its merits and in terms of its impacts on historically marginalized groups and society at large.

Considerations such as the purpose of an AI tool or system, the implicit framing of a problem to which the technology is proposed as the solution, and the need for transparency about funders of a particular project should thus receive early emphasis similar to that given the "technical characteristics" (which themselves are not all necessarily only technical) associated with "trust" listed on page 1 (lines 199-200).

Moreover, different groups and individuals in society will find it easier or harder to "trust" algorithmic tools deployed by major corporations or by the state, depending on their prior interactions with either or both, in addition to historical and sociopolitical context. The acceptability of an algorithmic system should depend on actual impacts, rather than on the opinions of people who may belong to social groups privileged by that particular system, and who may benefit from it but at the expense of historically marginalized and vulnerable communities.

---

29 See the White Collar Crime Risk Zones project as further demonstration of this point: Brian Clifton, Sam Lavigne & Francis Tseng, *Predicting Financial Crime: Augmenting the Predictive Policing Arsenal*, The New Inquiry (April 2017), https://whitecollar.thenewinquiry.com/static/whitepaper.pdf; and Sam Lavigne, Francis Tseng & Brian Clifton, *White Collar Crime Risk Zones*, The New Inquiry (April 26, 2017), https://thenewinquiry.com/white-collar-crime-risk-zones/.

30 See e.g., Amber M Hamilton, "Silicon Valley Pretends That Algorithmic Bias Is Accidental. It's Not," *Slate* (July 7, 2021), https://slate.com/technology/2021/07/silicon-valley-algorithmic-bias-structural-racism.html.

31 Joan E Solsman, "Deepfake bot on Telegram is violating women by forging nudes from regular pics," *CNET* (October 22, 2020), https://www.cnet.com/tech/services-and-software/deepfake-bot-on-telegram-is-violating-women-by-forging-nudes-from-regular-pics/.

Recommendation 3: Incorporate analysis and recommendations from privacy law scholarship regarding the dangers of "managerialization" and legal endogeneity of privacy compliance, and apply them to the algorithmic bias context.

The draft proposal's repeated references to and framing of "managing bias" risks establishing a regulatory structure that amounts to "managerialization" of algorithmic bias rather than meaningful prevention of it. The term "managerialization" describes a situation where "symbolic structures of compliance" with, for example, privacy law, such as paper trails and procedural checkboxes, are established to serve corporate risk management purposes, but over time are upheld as the law itself in operation, even if no privacy rights are being upheld and no one is in reality protected.[32] This can lead to a similar and related phenomenon, "legal endogeneity", which is when a legal regime "elevat[es] form over substance, catalyzing the development of compliance structures that, on their face, seem to comply with the law, but, as mere symbols of compliance" in fact defeat the purpose of protecting impacted groups and individuals.[33]

While the concepts of managerialization and legal endogeneity are taken from the privacy law context, Ari Waldman's analysis and recommendations can also apply to the work NIST is doing to address bias in AI. The potential danger in an approach based on "*managing*" bias — as opposed to, for instance, preventing it or providing legal recourse for those impacted — is a resulting scheme of AI vendors and clients, aided by a complementary industry of compliance or "AI bias" consultants, going through the motions of completing technical standards checklists, which then counts as legal compliance, while continuing to violate the substantive legal rights of historically marginalized groups regardless.[34] To increase chances of avoiding such an outcome, NIST should ensure that its proposed approach will not enable harmful or violative practices involving algorithmic bias to continue simply because the relevant entities have executed compliance processes.

Recommendation 4: Add a fourth stage to the "AI lifecycle": post-deployment, which should have its own set of obligations distinct from those in the active deployment stage and its immediate aftermath.

NIST may wish to consider adding a fourth stage to the "AI lifecycle": post-deployment. The post-deployment phase would begin after the immediate aftermath of the active deployment stage of an algorithmic tool or system, and would last until the tool or system is no longer in operation. A distinct set of obligations should attach to the full duration of the post-deployment phase, including routine audits, testing, and evaluation; updated algorithmic impact

---

32 Ari Ezra Waldman, *Privacy Law's False Promise*, 97:3 Washington University Law Review 773 (2020) at 777-78, https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=6386&context=law_lawreview.

33 *Ibid.*, at 825.

34 See e.g., in the context of privacy law, Julie E. Cohen, *How (Not) to Write a Privacy Law*, Knight First Amendment Institute (March 23, 2021), https://knightcolumbia.org/content/how-not-to-write-a-privacy-law; Ari Ezra Waldman, *Privacy Law's False Promise*, 97:3 Washington University Law Review 773 (2020), https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=6386&context=law_lawreview; and Ari Ezra Waldman, *Privacy, Practice, and Performance*, 110 California Law Review (forthcoming, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3784667.

assessments; complaint or recall mechanisms; and/or public justification of why the system should remain in operation.

Structuring the lifecycle to include this fourth stage and the above obligations would counter the impulse to consider the process "done and over with" once the algorithm has been deployed, and encourage sustained engagement with its potential impacts as long as it exists. Too often, impacted historically marginalized and vulnerable communities are left to contend with the consequences of harmful algorithmic decision-making systems on their own, while those who deployed or benefit from it have moved on. An explicit post-deployment stage with attendant obligations may mitigate or prevent continuation of this kind of exploitative cycle.

Thank you for the opportunity to provide these comments.

Respectfully submitted,

/s/__________________

Cynthia Khoo
Center on Privacy & Technology
Georgetown Law
600 New Jersey Avenue, NW
Washington, DC 20001
(202) 662-9000