SURVEILLANCE TECHNOLOGY OVERSIGHT PROJECT

40 Rector Street, 9th Floor
New York, New York 10006
www.StopSpying.org | (646) 602-5600

COMMENT OF THE
SURVEILLANCE TECHNOLOGY OVERSIGHT PROJECT

TO THE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

IN RESPONSE TO
FIRST PUBLIC DRAFT OF NIST SP 1270: "A PROPOSAL FOR IDENTIFYING AND MANAGING BIAS IN ARTIFICIAL INTELLIGENCE"

SUBMITTED
SEPTEMBER 10TH, 2021

We thank NIST for inviting comments on its draft report, "A Proposal for Identifying and Managing Bias in Artificial Intelligence." Our comments align with the authors' three stages of managing AI bias. To summarize:

- During the pre-design phase, NIST should categorically oppose many of the AI tools under consideration. NIST underestimates the degree to which tools' harms can be anticipated or prevented, particularly in areas like policing, where both errors and the accurate use of AI can have devastating impacts.
- During the design and development phase, we recommend greater humility regarding the degree to which algorithmic tools can be debiased. Algorithms trained on police administrative data incorporate historical patterns of police abuse and bias. "Debiasing" may indemnify developers against liability for bias they fail to meaningfully curb.
- During the deployment stage, we recommend that NIST acknowledge the rights violations that occur when police AI is misused (regardless of whether tools are "debiased").

NIST's authors suggest that "[i]nstead of viewing the challenge of AI bias within a given context or use case… [we] strike the problem of AI bias where it might be easiest to manage – within the design, development, and use of AI systems." NIST's generalized approach glosses over the unique harms AI threatens in sectors like policing and criminal justice, underemphasizing our obligation to protect the public from systems that not only can increase injustice, but threaten Americans lives. NIST states that "[t]he goal is not zero risk but rather, identifying, understanding, measuring, managing and reducing bias." To the contrary, police AI is inherently incompatible with the public's safety, liberty, and fundamental rights. We must not merely mitigate policing tools' risk, but instead truly protect the public by banning tools that impose too great a cost on society.

I. The Pre-Design Phase: Focus on whether a tool should be built at all

Police AI exacts such a predictable toll on civil rights that these tools should have been blocked during NIST's "pre-design" phase. And such systems would have been blocked if developers seriously evaluated their social impact. But given vendors' eagerness to sell policing technology, NIST is dangerously dismissive of *the* key pre-design question: should we build a tool at all? Or rather, because police AI typically is used first in other fields, should an existing technology be imported into policing?

The question of whether to adapt algorithmic tools for law enforcement use could not be more momentous. Individuals' freedom from wrongful imprisonment, safety, privacy, freedom of association and other fundamental rights have been jeopardized by the importation of AI systems into policing. Predictive policing algorithms have justified the continued, dangerous over-

policing of BIPOC neighborhoods by police precincts with a history of racial bias.[1] Facial recognition errors have already been documented as causing the wrongful arrest of several Black men, though many more individuals have likely been impacted by the technology without knowing.[2] ShotSpotter errors bring armed police into Black and Latinx communities under conditions primed for deadly mistakes.[3] Criminal justice algorithms routinely mete out biased recommendations for pretrial detention and imprisonment of Black and Latinx individuals.[4]

Any adequate pre-design phase would require developers to demonstrate that they are not replicating the same sort of deadly errors showcased by this technology to date. The tools described above could have been and should have been abandoned during development. And those of us who work in the police technology space have seen enough such tools to anticipate the civil rights violations that future will introduce.

Consider the devastating and foreseeable effects of the New York City Police Department's ("NYPD's") use of PredPol beginning in 2013. PredPol is an algorithmic tool that claims to predict where and when crimes will occur and who will commit it.[5] PredPol predictably focused police on the low-income BIPOC communities targeted by NYPD officers for years, with devastating results.[6]

Had PredPol's developers considered the probable effects of the technology during the pre-design phase, the following facts would have stood out:

I. Police discrimination against BIPOC communities has distorted historical policing data.[7]
II. Police encounters are disproportionately dangerous for these same communities. Black men are two and a half times more likely than white men to be fatally shot by police,[8] and they are the victims in one of three fatal traffic stops.[9]

---

[1] Rashida Richardson, Jason Schultz, and Kate Crawford, "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice," *N.Y.U. Law Review Online* 94, no. 192 (February 13, 2019), https://papers.ssrn.com/abstract=3333423.
[2] Kashmir Hill, "Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match," *The New York Times*, December 29, 2020, sec. Technology, https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html.
[3] "Comments on Draft NYPD Surveillance Policies," Center for Constitutional Rights, February 25, 2021, https://ccrjustice.org/node/9092.
[4] Julia Angwin et al., "Machine Bias," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=m1ze0Mrj6m52j-J8AvluRGJmCGvDt8BG.
[5] Tim Lau, "Predictive Policing Explained" (Brennan Center for Justice, April 1, 2020), https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained.
[6] Rashida Richardson, "Dirty Data, Bad Predictions."
[7] Josmar Trujillo, "Why NYPD's 'Predictive Policing' Should Scare You," *City Limits*, January 29, 2015, sec. CITY VIEWS: OPINIONS and ANALYSIS, https://citylimits.org/2015/01/29/why-nypds-predictive-policing-should-scare-you/.
[8] Lynne Peeples, "What the Data Say about Police Brutality and Racial Bias — and Which Reforms Might Work," *Nature* 583, no. 7814 (June 19, 2020): 22–24, https://doi.org/10.1038/d41586-020-01846-z.
[9] Wesley Lowery, "A Disproportionate Number of Black Victims in Fatal Traffic Stops," *Washington Post*, December 24, 2015, sec. National, https://www.washingtonpost.com/national/a-disproportionate-number-of-black-victims-in-fatal-traffic-stops/2015/12/24/c29717e2-a344-11e5-9c4e-be37f66848bb_story.html.

III. Users use products to "tech-wash" biased behavior.[10] Developers could anticipate that officers would use PredPol to justify biased policing and avoid scrutiny.[11]

In short, developers could have foreseen the very reasons not to move forward with building predictive policing tools: they perpetuate racist policing, subjecting BIPOC communities to dangerous, continual police harassment and ensnaring individuals in the criminal justice system.

NIST's authors do acknowledge the kinds of reasons that could lead to a tool's abandonment in the pre-design phase:

> It is an obvious risk to build algorithmic-based decision tools for settings already known to be discriminatory.
>
> [P]re-design is often where decisions are made that can inadvertently lead to harmful impact, or be employed to extremely negative societal ends.

But NIST plays the apologist, downplaying developers' ability to foresee civil rights concerns:

> [A]wareness of which conditions will lead to disparate impact or other negative outcomes is not always apparent in pre-design, and can be easily overlooked once in production.

Instead, NIST focuses on managing risk while moving risky projects forward:

> [W]ell-developed guidance, assurance, and governance processes can assist business units and data scientists to collaboratively integrate processes that reduce bias without being cumbersome or blocking progress.

And NIST casts the decision to stop tools' development as a rare, "extreme" measure rather than a reasonable and common solution:

> In extreme cases, with tools or apps that are fraudulent, pseudoscientific, prey on the user, or generally exaggerate claims, the goal should not be to ensure tools are bias-free, but to reject the development outright.

Contrary to NIST's contention, pseudoscience and exaggerated claims are not "extreme cases"; they are typical for policing AI. It is not enough, having anticipated "extremely negative societal ends," to ensure that "risk management processes… set reasonable limits related to mitigating such potential harms." Contrary to NIST's contention, effective limits frequently cannot "reduce bias without being cumbersome or blocking progress." The correct response to biased and invasive tools is to simply stop their development and sale completely.

This recommendation is not extreme. Other scientific disciplines long recognized that some advances simply come at too high a price. Consider the breakthroughs that scientists could have made in chemical and biological warfare over the past 50 years if permitted. Such agents would be potent weapons in our military arsenal, but they would pose an intolerable risk to all of humanity. Many of the AI systems under development—and indeed, many in use today—pose

[10] John D. Lee and Katrina A. See, "Trust in Automation: Designing for Appropriate Reliance," *Human Factors* 46, no. 1 (2004): 50–80, https://doi.org/10.1518/hfes.46.1.50_30392.
[11] Josmar Trujillo, "NYPD's 'Predictive Policing'."

an intolerable cost to human rights and civil rights, both here in the United States and when exported abroad. Before we ask how to build better AI, we must ask if that AI is truly an acceptable solution for the problems we purport to solve.

II. The Limits of Debiasing in the Design and Development Phase

NIST states that algorithmic tool bias can be mitigated during the design and development phase:

> Instead of viewing the challenge of AI bias within a given context or use case, a broader perspective can strike the problem of AI bias where it might be easiest to manage – within the design, development, and use of AI systems.
>
> These unintentional weightings of certain factors can cause algorithmic results that exacerbate and reinforce societal inequities. The surfacing of these inequities is a kind of positive "side effect" of algorithmic modeling, enabling the research community to discover them and develop methods for managing them.

This is a dangerously idealistic approach to ending technology-aided discrimination, particularly in fields like policing. If an algorithm does not "exacerbate and reinforce social inequities" in a lab, it easily may do so in the real world.

The difficulty with debiasing policing algorithms has to do with how the tools become biased in the first place:

> Historical, training data, and measurement biases are "baked-in" to the data used in the algorithmic models underlying those types of decisions. Such biases may produce unjust outcomes for racial and ethnic minorities in areas such as criminal justice.

Policing algorithms' "training data"—the data that models correct behavior for algorithms—is administrative data that police departments and courts collect on a day-in, day-out basis. Police records includes information on reported incidents, police stops, arrests and charges leading to arrest. Court records add information on convictions and acquittals, pretrial detention and bail, and other details about individuals' passage through the criminal justice system. But those records include the results of biased, corrupt, and criminal policing.[12]

Here in New York City, NYPD records memorialize unconstitutional practices such a Stop-And-Frisk, which targeted 5 million individuals who were stopped from 2002-2013 in a practice likened to a police "war with Black and Brown people."[13] In 2013, a federal court determined

[12] Rashida Richardson, "Dirty Data, Bad Predictions."
[13] Ashley Southall and Michael Gold, "Why 'Stop-and-Frisk' Inflamed Black and Hispanic Neighborhoods," *The New York Times*, November 17, 2019, sec. New York, https://www.nytimes.com/2019/11/17/nyregion/bloomberg-stop-and-frisk-new-york.html.

that stop-and-frisk violated the Fourth and 13th Amendments.[14] But algorithms trained on administrative data collected from 2002-2013 still learn how to replicate the NYPD's racial profiling from the height of stop-and-frisk.

NYPD records are also shaped by police criminality, including falsifying records, arbitrary arrest and summons quotas, and planting evidence on innocent New Yorkers.[15] In 2017 alone NYC taxpayers paid $335 million to victims of police abuse.[16] Those crimes—including many wrongful arrests—are "baked in" to whatever tools the NYPD trains. Even data that is supposed to have been expunged is still part of the NYPD's records. As of at least 2018, officers still had routine access to data on dropped, declined, and dismissed arrests that should have been expunged pursuant to state law.[17]

If we simply focus on fixing the algorithm, as NIST suggests, the technical solution is to seek out a more balanced training dataset—one that doesn't systematically target BIPOC communities. But there is no unbiased dataset for policing in America, just records bathed in bias, memorializing practices that no developer should seek to emulate.

Proponents of algorithmic policing tools suggest "cleaning" training data or balancing its outputs to reduce algorithms' disparate impacts. We believe that it is dubious that developers have the proper incentives to effectively implement such strategies—something that frequently may not be technically possible. Rather, they will use such techniques to minimize liability and reputational risks. It would be the height of "data hubris," to use NIST's term, to imagine that a debiased algorithm can correct centuries of systemic discrimination.

III. Bias and Misuse in the Deployment Phase

In its discussion of the deployment phase, NIST anticipates "off-road" uses of algorithms—unplanned uses where "the tool is used in unforeseen ways." This bland description does not capture the gravity of police technology abuses. As seen with the misuse of ShotSpotter and facial recognition technology, even if developers fixed technical drivers of algorithmic bias, such efforts would fail to address the true gravity of police AI's threat to the public.

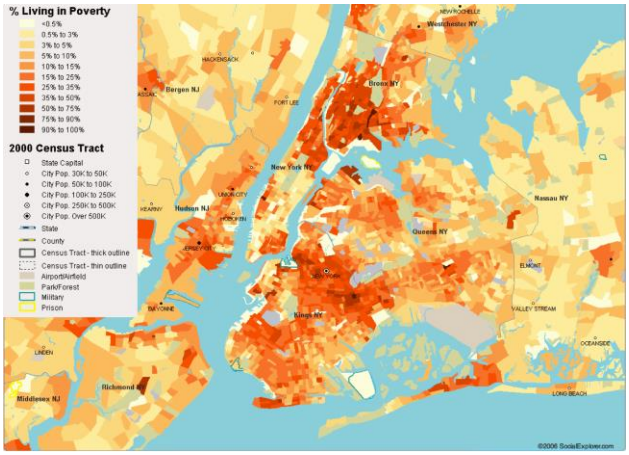
---

[14] Joseph Goldstein, "Judge Rejects New York's Stop-and-Frisk Policy," *The New York Times*, August 12, 2013, sec. New York, https://www.nytimes.com/2013/08/13/nyregion/stop-and-frisk-practice-violated-rights-judge-rules.html.
[15] Rashida Richardson, "Dirty Data, Bad Predictions."
[16] Jake Offenhartz, "Lawsuits Against NYPD Cost Taxpayers $230 Million Last Year," Gothamist, April 17, 2019, https://gothamist.com.
[17] Eli Hager, "Your Arrest Was Dismissed. But It's Still In A Police Database.," The Marshall Project, July 18, 2019, https://www.themarshallproject.org/2019/07/18/your-arrest-was-dismissed-but-it-s-still-in-a-police-database.

NYPD's Biased Placement of ShotSpotter Units

ShotSpotter is a for-profit corporation that markets systems that use audio surveillance and algorithmic software to purportedly locate gunshots. ShotSpotter boasts high accuracy levels in laboratory conditions,[18] though its real-world performance is wanting.[19] ShotSpotter's algorithm may not be biased, but it's placement overwhelmingly in low-income BIPOC communities is. As shown below, the yellow areas on the left where ShotSpotter is deployed in New York City largely mirror the red neighborhoods on the right with the highest levels of poverty.

Shotspotter deployments in 2018. Photo credit: Clare Garvey.



Map of Poverty in NYC. Credit: Visualizing Economics.

Since ShotSpotter is highly error prone, with one study finding that 89% of reports were false,[20] ShotSpotter's biased placement makes BIPOC communities bear the constant cost of armed officers rushing to the scene of shootings that never happened.[21] In one tragic example earlier this year, police responded to a ShotSpotter report of gunshots. [22] Five minutes later, they shot and killed Adam Toledo, a 13-year-old who was holding his empty hands up when he died.[23] ShotSpotter's bias can't be appreciated in a lab, but biased ShotSpotter deployment is already endangering real world neighborhoods in Chicago, New York, and countless other cities

[18] "ShotSpotter Repond Q&A," *ShotSpotter*, December 2020, https://www.shotspotter.com/wp-content/uploads/2020/12/ShotSpotter-Respond-FAQ-Dec-2020.pdf.
[19] *See*, for example, "End Police Surveillance," Roderick & Solange MacArthur Justice Center, 2021, https://endpolicesurveillance.com/.
[20] "End Police Surveillance."
[21] "End Police Surveillance: The Burden on Communities of Color," Roderick & Solage MacArthur Justice Center, 2021, https://endpolicesurveillance.com/.
[22] Timothy R. Homan, "Police Technology under Scrutiny Following Chicago Shooting," Text, TheHill, April 21, 2021, https://thehill.com/homenews/state-watch/549612-police-technology-under-scrutiny-following-chicago-shooting.
[23] Chrisoph Koettl and Evan Hill, "How an Officer Killed Adam Toledo: Video Investigation - The New York Times," *The New York Times*, April 14, 2021, https://www.nytimes.com/2021/04/16/us/adam-toledo-video-investigation.html.

NYPD Abuse of Facial Recognition

The NYPD uses at least two facial recognition ("FR") vendors: DataWorks Plus, its main vendor, and Clearview AI, which NYPD has used on an extended trial basis. The tools' performance in the lab is unknown. DataWorks does not conduct accuracy and bias testing, according to one of its own managers.[24] Clearview AI generally does not submit its system for outside testing and its accuracy and bias are not publicly known. Clearview's tool displays a disturbing lack of respect for privacy: it identifies unknown individuals by comparing their photos to "3 billion photos scraped from the web,"[25] forcing individuals in those billions of photos to stand in a "perpetual line-up."[26]

But the most disturbing thing about NYPD's FR tools is not how they perform in the lab—it is how the department uses them. NYPD officers are free to misuse and abuse the tools, exercising "artistic license" with photos to improve their chances of finding a supposed match, having an unmeasurable impact on accuracy and bias. According to a report on NYPD FR practices from Georgetown's Center on Privacy and Technology, photo "edits often go well beyond minor lighting adjustments and color correction, and often amount to fabricating completely new identity points not present in the original photo."[27] NYPD has replaced features and expressions in street-camera photos with features from mugshots.[28] It has used "3D modeling software to complete partial faces" and to "rotate faces that are turned away from the camera."[29] By scanning altered photographs, officers destroy what little credibility FR has as a reliable source of identification.[30] Police practices routinely transform the technology into the very sort of pseudoscience that NIST dismissed as "extreme cases."

Even worse, the NYPD primarily compares probe images against a gallery of historical mugshots, skewing the risk of false "matches" to disproportionately impact the BIPOC communities who have long faced higher arrest rates. Like use of NYPD data to train PredPol, this practice can compound the impact of biased police practices even years after they take place.[31]

---

[24] Kashmir Hill, "Wrongfully Accused by an Algorithm," *The New York Times*, June 24, 2020, sec. Technology, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.
[25] Tate Ryan-Mosley, "The NYPD Used Clearview's Controversial Facial Recognition Tool. Here's What You Need to Know," MIT Technology Review, April 9, 2021, https://www.technologyreview.com/2021/04/09/1022240/clearview-ai-nypd-emails/ .
[26] Clare Garvie, Alvaro Bedoya, and Jonathan Frankle, "The Perpetual Line-Up" (Center on Privacy & Technology at Georgetown Law, October 16, 2016), https://www.perpetuallineup.org/.
[27] Clare Garvie, "Garbage In, Garbage Out: Face Recognition on Flawed Data," Georgetown Law Center on Privacy and Technology, May 16, 2019, https://www.flawedfacedata.com.
[28] Garvie, "Garbage In, Garbage Out."
[29] Garvie, "Garbage In, Garbage Out."
[30] Garvie, "Garbage In, Garbage Out."
[31] Mariko Hirose, "Privacy in Public Spaces: The Reasonable Expectation of Privacy against the Dragnet Use of Facial Recognition Technology," *Connecticut Law Review* 49, no. 5 (September 2017): 1591–1620.

Lastly, when using FR, NYPD officers manually select the winning "match" from a list of hundreds of possible results. This adds yet another layer of bias—and potential for outright abuse—in the facial recognition decision process. NIST suggests that when a gap exists between an algorithm's intended use and its actual use, the solution may be "deployment monitoring and auditing" followed by adjustments to the algorithmic model. But there is no algorithmic fix that will correct officers' bias or misuse of AI, particularly facial recognition.

IV. Conclusion

As NIST finalizes its report, we recommend revisions that acknowledge the civil rights violations that AI policing tools enable. In the pre-design phase, developers must abandon tools that risk acute harms and high rates of bias. In the design and development phase, developers must acknowledge limits on technical debiasing, especially for policing tools. And during deployment, supposedly "debiased" tools must not be permitted to be deployed in biased ways. More broadly, NIST must look beyond technical fixes to algorithms. Only by addressing human behavior and systemic bias can we address the racism and injustice AI enables and augments.