INNOCENCE PROJECT

Executive Director
Christina Swarns, Esq.

Co-Founders & Special Counsel
Barry C. Scheck, Esq.
Peter J. Neufeld, Esq.

INNOCENCE PROJECT PUBLIC COMMENT ON
DRAFT NIST Special Publication 1270
A PROPOSAL FOR IDENTIFYING AND MANAGING BIAS WITHIN ARTIFICIAL INTELLIGENCE
September 10, 2021

The Innocence Project is pleased to respond to the National Institute of Standards and Technology (NIST) call for public comments regarding the Draft NIST Special Publication 1270, *A Proposal for Identifying and Managing Bias within Artificial Intelligence* ("the report"). For nearly 30 years, the Innocence Project has worked to exonerate the innocent and prevent wrongful convictions through systemic reform. In cases where we have proven innocence, misapplied forensic science contributed to 52% of the wrongful convictions.[1] The vast majority of our exonerations were achieved by the power and strength of forensic DNA evidence. However, we have watched with concern how—through technologies like Rapid DNA and familial searching—DNA applications have expanded beyond truth seeking instruments into tools of surveillance that target innocent people, exacerbate racial disparities, and promote the unsupported notion that criminality is genetic.[2] Based on these decades of experience, the Innocence Project takes the position that, in addition to meeting scientific metrics of validity and reliability, the research and development of criminal legal system applications must simultaneously assess social impact, considering ethical, legal, and social implications, and capacity for just and equitable implementation. Any framework for managing bias in AI systems must simultaneously address both the scientific underpinnings of the technology as well as social consequences.

A primary concern of the Innocence Project's comments on the proposed framework for managing AI bias ("the Framework") is how the Framework impacts suspect development. Blanket intelligence systems and surveillance technologies built on algorithms can entrap the innocent by creating an entry point to wrongful convictions.[3] Once an innocent person becomes a person of interest through the use of blanket intelligence systems and surveillance technologies, tunnel vision sets in,

[1] Innocence Project, *Overturning Wrongful Convictions Involving Misapplied Forensics*, INNOCENCE PROJECT, https://www.innocenceproject.org/overturning-wrongful-convictions-involving-flawed-forensics/ (last visited Sep 6, 2019).
[2] Erin E. Murphy, Inside the Cell: The Dark Side of Forensic DNA (2015); Erin Murphy, *Relative Doubt: Familial Searches of DNA Databases*, 109 Mich. Law Rev. 59 (2010); Nancy Gertner et al., *Report on S.2480, "An Act Permitting Familial Searching and Partial DNA Matches in Investigating Certain Unsolved Crimes" and Related Recommendations Pertaining to G.L. c.22E Governing the Massachusetts Statewide DNA Database* (2021).
[3] Rebecca Brown, *3 Ways Lack of Police Accountability Contributes to Wrongful Convictions*, INNOCENCE PROJECT (2020), https://innocenceproject.org/lack-of-police-accountability-contributes-to-wrongful-conviction/ (last visited Aug 30, 2021).

and no amount of exculpatory evidence can derail an investigator's conviction of the innocent person's guilt. Exonerations demonstrate this dynamic. Pre-trial exculpatory DNA results were explained away or dismissed in 28 of the 325 DNA exonerations in the United States between 1989-2014.[4]

Secondly, AI systems cannot be separated from the policing systems that administer them; their applications to society and the data these technologies collect will reflect the disparities, flaws, and biases of those law enforcement practices.[5] Racially disparate policing perpetually criminalizes communities of color and promotes false narratives that impact how these communities are perceived by law enforcement. For example, Rock Harmon, a former prosecutor and familial DNA testing advocate has repeatedly stated in different fora that "Familial DNA searching relies on the premise that crime runs in families."[6] This false and scientifically unsupported narrative conditions police to treat entire communities as trouble zones and contributes to racially disparate policing practices and mass incarceration.[7] Consequently, AI surveillance technologies are "suspect development systems" when the government uses them to "manage vague or often immeasurable social risks based on presumed or real social conditions" and "subjects targeted individuals or groups to greater suspicion, differential treatment, and more punitive and exclusionary outcomes."[8]

For these reasons, blanket surveillance or investigative systems used to develop suspects, such as gang databases, pose social risks—especially for groups of people who have historically been the target of surveillance. To narrow the entry point for innocent people into the a criminal legal system built on extracting convictions regardless of an individual's actual guilt, it is the Innocence Project's position that investigative technologies must meet the same standards of accuracy and reliability expected of court admissible evidence, and demonstrate their capacity for just and equitable implementation prior to their implementation in the criminal legal system.[9] To require anything less is tantamount to facilitating the experimentation of these technologies on society. This is a painful and intolerable risk. The narrative that policing strategies and due process will weed out

[4] Emily West & Vanessa Meterko, *Innocence Project: DNA Exonerations, 1989-2014: Review of Data and Findings from the First 25 Years*, 79 ALBANY LAW REV. 717–795 (2016).
[5] Rashida Richardson, Jason M Schultz & Kate Crawford, *DIRTY DATA, BAD PREDICTIONS: HOW CIVIL RIGHTS VIOLATIONS IMPACT POLICE DATA, PREDICTIVE POLICING SYSTEMS, AND JUSTICE*, 94 N. Y. UNIV. LAW REV. 42.
[6] Meredith Salisbury, *Are You Related to a Killer? Police Want to Know.*, TECHONOMY, 2019, https://techonomy.com/2019/05/are-you-related-to-a-killer-police-want-to-know/ (last visited Dec 13, 2020).
[7] Anthony A. Braga, Rod K. Brunson & Kevin M. Drakulich, *Race, Place, and Effective Policing*, 45 ANNU. REV. SOCIOL. 535–555 (2019); Elizabeth Hinton & DeAnza Cook, *The Mass Criminalization of Black Americans: A Historical Overview*, 4 ANNU. REV. CRIMINOL. null (2021).
[8] Rashida Richardson & Amba Kak, *Suspect Development Systems: Databasing Marginality and Enforcing Discipline*, 55 UNIV. MICH. J. LAW REFORM (forthcoming), https://www.ssrn.com/abstract=3868392 (last visited Jul 8, 2021).
[9] NATIONAL ASSOCIATION OF CRIMINAL DEFENSE LAWYERS, *The Trial Penalty: The Sixth Amendment Right to Trial on the Verge of Extinction and How to Save It* 331–368 (2019), https://online.ucpress.edu/fsr/article/31/4-5/331/109303/The-Trial-Penalty-The-Sixth-Amendment-Right-to (last visited Aug 11, 2021).

innocent people prior to conviction has been disproven by numerous wrongful convictions. That narrative also dismisses the seriousness and harm of collateral consequences of arrests. There is no dispute that Michael Oliver, Robert Williams, and Njeer Parks' wrongful arrests were the byproduct of both a flawed facial recognition system as well as flawed policing.[10] But for the fact these men held tightly to their innocence and their unjust arrests were recognized, they could have been railroaded into wrongful convictions. At this time, we cannot know the scope of people whose wrongful arrests were predicated on these technologies and the fact that Mr. Oliver, Mr. Williams, and Mr. Parks were eventually able to demonstrate their unjust arrests should provide no comfort that these errors can be comprehensively surfaced.

We take the time to share these concerns to emphasize a critical point—no amount of validation testing, standards development, or technical solutions will ensure the just and equitable application of AI technologies in the American policing system. While the development of a framework is an important first step to raising awareness regarding the harms that AI technologies can impose on society, the application of the framework will always be limited when the social, political, economic, and structural solutions required for justice are out of the scope of this proposal.

Thank you in advance for your consideration of the feedback we respectfully offer. Please find our comments below.

Public Comments

Our comments are integrated into the chart below, which is a modified version of the suggested template. When new language or edits are suggested to resolve comments regarding excerpts of the report, ~~strikethroughs~~ are used to indicate text that should be deleted and **[bracketed and bold text]** indicate text that should be added.

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
1	242-243	2	While it's unlikely that technology exhibiting "zero risk" can be developed, managing and reducing the impacts of harmful biases in AI is possible and necessary.	This sentence seems to suggest that while AI technologies may not be able to attain "zero risk" of harmful biases that it can get close. There is a vast universe of	This passage should be edited to indicate the limitations of the Framework:

[10] Kashmir Hill, *Wrongfully Accused by an Algorithm*, THE NEW YORK TIMES, June 24, 2020, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (last visited Jun 25, 2020); Elisha Anderson, *Controversial Detroit facial recognition got him arrested for a crime he didn't commit*, DETROIT FREE PRESS, July 10, 2020, https://www.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002/ (last visited Oct 26, 2020); Kashmir Hill, *Flawed Facial Recognition Leads To Arrest and Jail for New Jersey Man - The New York Times*, NEW YORK TIMES, December 29, 2020, https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html (last visited Apr 10, 2021).

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
				social, political, economic, and structural influences on AI bias that cannot be mitigated without policy changes that are outside the scope of the technical solutions that NIST suggests.	While it's unlikely that technology exhibiting "zero risk" can be developed, managing and reducing the impacts of harmful biases in AI ~~is possible and necessary~~ **[requires collaboration from all stakeholders, especially vendors and users. However, ensuring the just and equitable implementation of AI technologies necessarily requires social, political, economic, and structural policy changes that are out of the scope of the current document]**.
2	281-283	2	AI development teams often use proxies. For example, for "criminality," a measurable index, or construct, might be created from other information, such as past arrests, age, and region.	This example is frequently used, and algorithm developers favor the use of arrests as a metric because the data are readily available. However, arrests are a measure of police activity and enforcement productivity and do not serve as a measure of criminality (Sparrow, 2015). This is a harmful and misapplied measure that is skewed by the overpolicing of communities of color (Gaston, 2019) and should not be perpetuated. References: Gaston, S. (2019). Producing race disparities: A study of drug arrests across place and race*. Criminology, 57(3), 1–28. https://doi.org/10.1111/1745-9125.12207 Sparrow, M. (2015). *Measuring Performance in a Modern Police*	Please delete this example and replace with an example in which the proxy is an accurate and reasonable measure of the phenomenon.

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
				Organization (NCJ 248476). U.S. Department of Justice, National Institute of Justice.	
3	320-321 323-324 328-330	2	Often a technology is not tested – or not tested extensively – before deployment, and instead deployment may be used as testing for the technology. There are also examples from the literature which describe technology that is based on questionable concepts, deceptive or unproven practices, or lacking theoretical underpinnings [2,9,13,30,33,62,129,141]. The decisions based on these algorithms affect people's lives in significant ways, and it is appropriate to expect protections in place to safeguard from certain systems and practices.	These passages are important for conveying the fact that society is a testing ground for many technologies and have harmful effects. The passages, however, do not communicate that technologies applied in the criminal legal system not only impact life and liberty, but also that the vast majority of jurisdictions implement these technologies without any regulatory framework or system in place to address the harms they produce.	Please add the following edits to lines (328-330): The decisions based on these algorithms affect people's lives in significant ways **[and can jeopardize life and liberty.]**~~, and it~~ **[It]** is appropriate to expect protections **[be put]** in place to safeguard ~~from~~ certain systems and practices. **[Currently, technologies in the criminal process are deployed in most jurisdictions in the absence of regulatory oversight and without systems in place to address the harms they produce.]**
4	347-354	2	Improving trust in AI systems can be advanced by putting mechanisms in place to reduce harmful bias in both deployed systems and in-production technology. Such mechanisms will require features such as a common vocabulary, clear and specific principles and governance approaches, and strategies for assurance. For the most part, the standards for these mechanisms and associated performance measurements still need to be created or adapted. The goal is not "zero risk," but to manage and reduce bias in a way that contributes to more equitable outcomes that engender public trust. These challenges are intertwined in complex ways and are unlikely to be addressed with a singular focus on	This passage exhorts standards and performance measures as the pathway to managing and reducing bias in such a way that it contributes to more equitable outcomes. However, there will be instances in which it is impossible to manage bias. For example, the selection of inappropriate performance measures (such as the use of past arrests as a proxy for criminality) or when the data collected is the product of biased activity.	Please add the following at the end of the paragraph at line (354): **[For example, technical approaches for applications in policing may meet insurmountable barriers to equitable bias mitigation. The implementation of AI technologies in this setting are applied in the context of policing activity and generates data that is inherently biased from its inception.]**

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
			one factor or within a specific use or industry.		
5	432-438	4	Central to these decisions is who (individuals or groups) makes them and which individuals or teams have the most power or control over them. These early decisions and who makes them can reflect individual and group heuristics and limited points of view, affect later stages and decisions in complex ways, and lead to biased outcomes [12,31,43,72,109,120]. This is a key juncture where well-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.	There is a reference to working with "new stakeholders" in line (375) and this passage misses the opportunity to name the parties that need to be at the table to establish the principle of inclusive participation. Inclusive participation in the development of technology improves public confidence and is key to encouraging the development of a shared morality that is essential to establishing the technology's legitimacy (Dryzek et al., 2020). Marginalized perspectives (Dryzek et al., 2020) and representatives of constituencies most impacted by the decision (Krimsky, 1984) are essential to producing more just decisions, based on richer knowledge resources, and ultimately improve public confidence (Krimsky, 1984). As Krimsky states, "To achieve the full benefits of participation by citizens, early access routes to the decision-making process should be developed, possibly even during the stage at which the problem is defined." <u>References:</u> Dryzek, J. S., Nicol, D., Niemeyer, S., Pemberton, S., Curato, N., Batterham, P., Bedsted, B., Burall, S., Burgess, M., Burgio, G., Castelfranchi, Y., Chneiweiss, H., Church, G., Crossley, M., de Vries, J., Farooque, M., Hammond, M., He, B., Mendonça, R., ... Rasko, J. E. J. (2020). Global citizen deliberation on genome editing. Science, 369(6510), 4. Krimsky, S. (1984). Beyond	Please add the following at the end of the sentence in line (433): Central to these decisions is who (individuals or groups) makes them and which individuals or teams have the most power or control over them. **[Inclusive participation will be key to the success of this framework and marginalized perspectives as well as the constituencies most impacted by the outcomes of the technology need to be integrated and empowered in decisionmaking at this stage of the framework.]**

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
				Technocracy: New Routes for Citizen Involvement in Social Risk Assessment. In J. C. Petersen (Ed.), Citizen Participation in Science Policy. The University of Massachusetts Press.	
6	447-450	4	It is an obvious risk to build algorithmic-based decision tools for settings already known to be discriminatory. Yet, awareness 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.	This statement calls into question AI technologies deployed in the criminal process. It needs to be highlighted and strengthened in the report because it raises significant red flags.	Please edit the statement to the following: **[There is a serious]** ~~It is an obvious~~ risk **[to society]** to build algorithmic-based decision tools for settings already known to be discriminatory**[, such as the criminal legal process,]**~~. Yet,~~ **[because]** awareness 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. **[Implementing these technologies without guidelines in place to establish restraints, justice and equity metrics to measure their performance, and systems in place to track implementation and correct and remedy harms they may produce, can generate serious harms to life and liberty.]**
7	457-460	4	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	The report vehemently rejects the use of "fraudulent, pseudoscientific" tools that "prey on the user," however harms can arise despite the best intentions of developers, stakeholders, and	Please edit the statement to the following: **[When tools or apps cause harm, whether they are]** ~~In extreme~~

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
			outright in order to prevent disappointment or harm to the user as well as to the reputation of the provider.	users. By only naming harm as the consequence of intentional misconduct, the Framework does not sufficiently address problems that are replicated by biased technologies which may be invisible to the people evaluating them. It also ignores the fact that there are market influences between vendors and police that are not based purely on the desire to create accurate technology. Harms have been documented in the use of inaccurate gang database algorithms (Howell and Bustamante, 2019; Speri, 2019) and facial recognition systems (Gilbert, 2020; Anderson, 2020; Hill, 2020; Hill, 2021), but their continued use belie that assumption. After all the sunk costs, developers may not be willing to rescind an AI tool after the pre-design stage if it resulted in harm and it would be incumbent upon stakeholders and users to reject it. References: Anderson, E. (2020, July 10). Controversial Detroit facial recognition got him arrested for a crime he didn't commit. Detroit Free Press. https://www.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002/ Gilbert, B. (2020, June 30). Facial-recognition software fails to correctly identify people "96% of the time," Detroit police chief says. Business Insider. https://www.businessinsider.com/facial-recognition-fails-96-of-the-time-detroit-police-chief-2020-6	~~cases, with tools or apps that are~~ fraudulent, pseudoscientific, prey on the user, or generally exaggerate claims, **[or if they are inadvertently designed to replicate disparities in society,]** the goal should not be to ensure tools are bias-free, but to reject the development outright in order to prevent disappointment or harm to the user as well as to the reputation of the provider. **[Users and stakeholders must be clear and steadfast in rejecting technologies that cause harm.]**

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
				Hill, K. (2020, June 24). Wrongfully Accused by an Algorithm. The New York Times. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html Hill, K. (2021, January 6). Flawed Facial Recognition Leads To Arrest and Jail for New Jersey Man—The New York Times. New York Times. https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html Howell, B., & Bustamante, P. (2019). Report on the Bronx 120 Mass "Gang" Prosecution. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3406106 Speri, A. (2019, April 25). The Largest Gang Raid in NYC History Swept Up Dozens of Young People Who Weren't In Gangs. The Intercept. https://theintercept.com/2019/04/25/bronx-120-report-mass-gang-prosecution-rico/	
8	462-466	4	Other problems that can occur in pre-design include poor problem framing, basing technology on spurious correlations from data-driven approaches, failing to establish appropriate underlying causal mechanisms, or generally technically flawed [22,34,40,52,54,89,102,110]. In such cases (often termed "fire, ready, aim"), the solution may not be mitigation, but rather, rejection of the system or the way in which the perceived underlying problem is framed.	This is a very important passage and the option to reject a technology must be on the table in the pre-design stage.	Please add the following at the end of the sentence in line (466): In such cases (often termed "fire, ready, aim"), the solution may not be mitigation, but rather, rejection of the system or the way in which the perceived underlying problem is framed. **[It is critical to note that rejecting the system or the technology must be**

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
					among the options on the table in this phase.]
9	466-468	4	These types of scenarios may reinforce public distrust of AI technology as systems that are untested or technically flawed can also contribute to bias.	It is not simply that harmful technologies reduce public trust because they contribute to bias. Harmful technologies breed and cultivate distrust because they cause significant harm and are essentially tested on communities of people with historical legacies of being the subject of unethical and abusive testing.	Please edit the statement to the following: These types of scenarios may reinforce public distrust of AI technology as systems that are untested or technically flawed ~~can also contribute to bias~~ **[are experimented upon society and often produce harm on communities of people with historical legacies of being the subject of unethical and abusive testing]**.
10	476-479	4	It is also complicated by the role of power and decision making [96]. A consistent theme from the literature is the benefit of engaging a variety of stakeholders and maintaining diversity along social lines where bias is a concern (racial diversity, gender diversity, age diversity, diversity of physical ability) [32]. These kinds of practices can lead to a more thorough evaluation of the broad societal impacts of technology-based tools across the three stages.	Integrating a diverse set of stakeholders who represent the diversity of perspectives as well as the constituencies most impacted by the technology is essential to the development of trustworthy AI. It is also notable that the various phases are not equally distributed in terms of time commitment. It would seem that the pre-design stage is fundamental to managing AI bias and should require the greatest investment of time.	Please edit the statement to the following: It is also complicated by the role of power and decision making [96]. A consistent theme from the literature is the benefit of engaging a variety of stakeholders and maintaining diversity **[and representation]** along social lines where bias is a concern (racial diversity, gender diversity, age diversity, diversity of physical ability) **[and among the constituencies who will bear the most harm from the technology]** [32]. These kinds of practices ~~can lead~~ **[are essential]** to a more thorough evaluation of the broad societal impacts of technology-based tools

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
					across the three stages. **[The pre-design stage is foundational to mitigating AI bias and requires a significant investment of time.]**
11	512-515	4	This stage of the AI lifecycle is where modeling, engineering and validation take place. The stakeholders in this stage tend to include software designers, engineers, and data scientists who carry out risk management techniques in the form of algorithmic auditing and enhanced metrics for validation and evaluation.	In the *Optimization over context* section in the Design and Development Stage, the report discusses the need to apply context in order to select the models that minimize bias, raised concerns about the use of aggregated data to make predictions about individual behavior, and that "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" (Lines 531-532). However, the composition of stakeholders that would participate in this stage of the Framework, as described in lines (512-515), do not sufficiently reflect the diversity of people who should be integrated in this stage to ensure the deliberation of context and its impact.	Please edit the statement to the following: This stage of the AI lifecycle is where modeling, engineering and validation take place. The stakeholders in this stage tend to include software designers, engineers, and data scientists who carry out risk management techniques in the form of algorithmic auditing and enhanced metrics for validation and evaluation. **[Stakeholders who represent marginalized perspectives as well as the constituencies most impacted by the outcomes of the technology must also be included in this phase as they are the best equipped experts to interpret how context may impact different models and metrics and the disparities they may cause.]**
12	583-585 596-599	4	Since many AI-based tools can skip deployment to a specified expert end user, and are marketed to, and directly used by, the general public, the intended uses for a given tool are often quickly overcome by reality. Once people start to interact with an AI system, early design and development decisions that were	In the *Discriminatory impact*, *Intended context v. actual context*, and *Contextual gaps lead to performance gaps* sections in the Deployment Stage, the report lists different ways that an algorithm, once deployed, can go wrong. Subsequently, in the *Practical improvements* section in the Deployment Stage, the report	At the end of the *Practical improvements* section at line (664), please add the following language in a new paragraph: **[However, monitoring and auditing approaches are limited to identifying how the algorithm failed**

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
			poorly or incompletely specified or based on narrow perspectives can be exposed. This leaves the process vulnerable to additive biases that are either statistical in nature or related to human decision making and behavior [109].	recommends monitoring and auditing deployment to manage bias risks and suggests using "counterfactual fairness" to improve the algorithm. However, this section does not describe how to correct or remediate the harms that were created by the problematic algorithm and should integrate language regarding the Duty to Correct and Notify.	**to perform as intended. Counterfactual fairness is a strategy to improve upon those detected failures. None of these strategies correct or remediate the harm that algorithms can impart upon the people for whom they fail. In the criminal process, algorithm failures jeopardize a person's life and liberty. Institutions that implement AI tools must establish Duty to Correct and Notify policies. The duty to correct and notify is an ethical and professional obligation of criminal legal system stakeholders when an adverse event occurs. Upon the discovery of the adverse event, the duty to correct requires that the party deploying the algorithm identify the affected cases, determine the system-level root and cultural causes, and remedy and correct all instances of the problem. The duty to notify requires the party deploying the algorithm and a diversity of system stakeholders to initiate a publicly accountable process to notify all individuals impacted by the adverse event.]**
	607-608		The deployment stage also offers an interesting window into how perceptions and uses can differ based on the distance from the technology itself.		
	638-640		Once the AI tool is deployed and goes "off-road," the original intent, idea, or impact assessment that was identified in pre-design can drift as the tool is repurposed and/or used in unforeseen ways.		
	645-650		There are individual differences in how humans interpret AI model output. When system designers do not take these differences into consideration it can contribute to misinterpretation of that output [21]. When these differences are combined with the societal biases found in datasets and human cognitive biases such as automation complacency (which is particularly relevant in the deployment stage), where end users may unintentionally "offload" their decisions to the automated tool - this can cause significant negative impacts.		
	654-658		One approach for managing bias risks associated with the gaps described above is deployment monitoring and auditing. Counterfactual fairness is a technique used by researchers to bridge the gaps between the laboratory and the post-deployment real world. The issue, as described in [81] is that "If		

Comment #	Paper Line #	Paper Section	NIST SP1270-DRAFT Language	Comments	Suggested Change(s)
			individuals in the training data have not already had equal opportunity, algorithms enforcing EO[6] will not remedy such unfairness."		