INNOCENCE PROJECT

**Executive Director**
Christina Swarns, Esq.

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



# INNOCENCE PROJECT PUBLIC COMMENT ON NIST AI Risk Management Framework Second Draft October 14, 2022

The Innocence Project is pleased to respond to the National Institute of Standards and Technology (NIST) call for public comments regarding the NIST *Artificial Intelligence Risk Management Framework: Second Draft* (AI RMF). For 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 risk management in AI systems must simultaneously address both the scientific underpinnings of the technology as well as the social consequences.

## AI Technologies Increase the Risk for Wrongful Conviction

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 ensnare 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, research demonstrates that tunnel vision sets in, and oftentimes not even powerful exculpatory evidence can derail an investigator's conviction of the innocent person's guilt. Exonerations demonstrate

[1] Innocence Project, Overturning Wrongful Convictions Involving Misapplied Forensics, Innocence Project, https://www.innocenceproject.org/overturning-wrongful-convictions-involving-flawed-forensics/ (last visited October 9, 2022).
[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); Dorothy Roberts, Fatal Invention (2011).
[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); 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).

this dynamic. This kind of investigatory tunnel vision has serious real world implications. For example, 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]

**AI Technologies Compound the Risk of Racially Disparate Policing**

Secondly, AI systems cannot be separated from the policing systems that administer them; how they are used and what data is collected 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, a prosecutor who advocated for familial DNA testing has repeatedly advocated 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] Because AI surveillance technologies extract and database location, biometric, and identity information, they become "suspect development systems" when they single out targeted or vulnerable groups for differential law enforcement treatment."[8]

For these reasons, blanket surveillance or investigative systems, such as gang databases, can sweep innocent people into the criminal legal system and increase their risk for wrongful conviction—especially for groups that have historically been the targets of surveillance. Given this high risk, the Innocence Project believes that investigative technologies must meet the same standards of accuracy and reliability expected of court-admissible evidence, and must demonstrate their capacity for just and equitable implementation before they are utilized in the criminal legal system.[9] To require anything less is tantamount to facilitating experimentation with these technologies on vulnerable segments of society, which would be a painful and intolerable risk. The contention that policing strategies and due process will weed out 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] At this time, we cannot know the

---

[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 (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. (2019).
[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 (2019); Elizabeth Hinton & DeAnza Cook, *The Mass Criminalization of Black Americans: A Historical Overview*, 4 Annu. Rev. Criminol. null (2021).
[8] Richardson and Kak, *supra* note 3.
[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 (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).
[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); 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); 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).

scope of people whose wrongful arrests were predicated on these technologies. 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, nor that the harms that such errors cause are justified.

**The NIST AI RMF Second Draft is an Important Start**

The NIST AI RMF is an important organizing and guidance document for the developers of AI technologies and those who implement them in the criminal legal system. While the authors have cautioned that the AI RMF is intended for voluntary adoption and designed at a high level to facilitate application across a broad spectrum of sectors, the Innocence Project's comments are specific to the application of AI RMF in criminal investigations and prosecutions. In these contexts, AI and automated technology have implications for constitutional rights and direct consequences on life and liberty. In our comments below, we will focus on the importance of the sociotechnical approach, the integration of stakeholders, ensuring transparency in criminal investigations and prosecutions, the duty to correct and notify, and how NIST can provide incentives for voluntary adoption of AI RMF.

## Recommendations

**Sociotechnical Approach to Ensure Justice and Equity**

The Innocence Project applauds NIST's recognition for the need to take a sociotechnical approach to evaluating the impact of AI and automated technologies on society, and its recognition of human factors and bias. These principles pierce the myth of objectivity that often surrounds the use of technology. Because AI technologies often use criminal legal system databases that have been populated through racially disparate policing, the implementation of these technologies cannot be objective and will replicate and exacerbate risks to overpoliced communities.[11] This process has been called "tech-washing," whereby surveillance tools are marketed and perceived as objective and race-blind, while they actually continue and further entrench historical racism.[12] This phenomenon is not new in the U.S. criminal legal system; it dates back to the use of post-Emancipation crime statistics.[13]

It is for this reason that we caution that the AI RMF Playbook's use of the status quo as a baseline for evaluating the context and frames related to AI systems deployed in the criminal legal system is a dangerous premise. The third step of the Map function in the AI RMF Playbook asks its users to gather information such that "AI capabilities, targeted usage, goals, and expected benefits and costs compared with the status quo are understood" (p. 22). Because most people in the United States live in jurisdictions where the criminal legal application of AI and automated technologies is unregulated and uncontrolled, the status quo is not only an extremely low bar, it also potentially uses unconstitutional practices as a baseline. But for the intervention of the courts in cases like *Carpenter v. United States* and *Leaders of a Beautiful Struggle v. Baltimore Police Department*,

[11] Richardson, Schultz, and Crawford, *supra* note 5; Richardson and Kak, *supra* note 3.
[12] National Association of Criminal Defense Lawyers, *Garbage In, Gospel Out*, (2021), https://www.nacdl.org/Document/GarbageInGospelOutDataDrivenPolicingTechnologies (last visited Sep 16, 2021).
[13] Khalil G. Muhammad, The Condemnation of Blackness: Race, Crime and the Making of Modern Urban America (2010).

a person's physical movements could be surveilled without a warrant by cell phone records and aerial surveillance.[14] While Fourth Amendment jurisprudence continues to evolve, we recommend that Map-3 be revised to require comparison not merely to "the status quo," which may reflect unconstitutional practices; but instead to "constitutional requirements." Otherwise, AI capabilities will simply replicate the unacceptable standards of current practice.

Lastly, the definitions for the fairness component of the AI RMF need to be better defined by its sociotechnical lens. By its own admission, the AI RMF states that "[s]tandards of fairness can be complex and difficult to define because perceptions of fairness differ among cultures and may shift depending on application. Systems in which biases are mitigated are not necessarily fair" (p. 14). A recent paper by Ben Green offers the concept of substantive fairness that is based on evaluations of both relational inequities (endemic in societal hierarchies) and structural inequities with regard to the implementation of algorithms.[15] Integrating the need to address relational and structural inequalities into the definition of fairness can respond to the limitations of more formal approaches to defining fairness.

**Recommendations:**

- Support the use of a sociotechnical approach to AI RMF and inclusion of language on human factors and biases.
- Edit Map-3 to: "AI capabilities, targeted usage, goals, and expected benefits and costs compared with ~~the~~ **constitutional requirements** are understood."
- Integrate the need to address relational and structural inequities in the definition of fairness in Section 4.3.

**Integration of Stakeholders**

The Innocence Project supports the thorough integration of stakeholders in both the NIST modification of the test, evaluation, verification, and validation (TEVV) cycle (Figure 1, p. 5) and the AI RMF Core (Figure 5, p. 17) and NIST's inclusion of affected communities among those defined as stakeholders. The stakeholders who have the power and opportunity to provide input in the criminal legal system are often those defined as its customers - law enforcement. With the exception of a handful of municipalities with citizen advisory boards, integrating the feedback and concerns of affected communities is not a routine practice in the criminal legal system. Ethicists have called for inclusive participation in the development of science and technology, and those who develop and implement AI and automated technologies should operate accordingly.[16]

Law enforcement deployment of technologies is often driven by availability of federal grant programs rather than based on an all-stakeholders analysis of the problem that the technology is intended to solve.[17] Impacted communities are essential to identifying and prioritizing their

[14] *Carpenter v. United States*, 138 S. Ct. 2206 (2018); *Leaders of a Beautiful Struggle v. Balt. Police Dep't*, 2 F.4th 330 (4th Cir. 2021).
[15] Ben Green, *Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness*, 35 Philos. Technol. 90 (2022).
[16] Sheldon Krimsky, *Beyond Technocracy: New Routes for Citizen Involvement in Social Risk Assessment*, *in* Citizen Participation in Science Policy (James C. Petersen ed., 1984).
[17] Samuel Nunn & Kenna Quinet, *Evaluating the Effects of Information Technology on Problem-Oriented-Policing: If it Doesn't Fit, Must We Quit?*, 26 Eval. Rev. 81 (2002).

endemic public safety challenges, and experts recommend using a problem-oriented framework.[18] The AI RMF offers recognition of this upstream decision by recommending that Framework users create "established processes for making go/no-go system commissioning and deployment decisions" (p. 16) and referencing the need to include stakeholders in this upstream in its first Mapping step (p. 21). However, the AI RMF does not explicitly connect the role of stakeholders influencing the go/no-go decision which is important for democratic policing processes and the use of surveillance technologies by public agencies.[19] Similarly, language should also explicitly connect the role of affected stakeholders in influencing AI RMF Manage phase decisions on whether to "supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use" (p. 25).

**Recommendations:**

- For the description of the TEVV cycle (Figure 1, p. 5) and the AI actors across the AI lifecycle (Figure 2, p. 6), include more explicit language in the text of Section 2 describing the need to integrate stakeholders throughout the cycle, both at the front end of the Plan & Design phase as well as the Operate & Monitor stage.
- In Section 6.2. Map, add a sentence after the first sentence of the second full paragraph on p. 21 to state: "After completing the Map function, Framework users should have sufficient contextual knowledge about AI system impacts to inform a go/no-go decision about whether to design, develop, or deploy an AI system based on an assessment of impacts. **It is especially important that stakeholders, especially affected communities, should have decision making powers when these technologies are deployed on behalf of the government upon its constituents."**
- In Section 6.4. Manage, add a sentence in the second paragraph on p. 25 to state: "After completing the Manage function, plans for prioritizing risk and continuous monitoring and improvement will be in place. Framework users will have enhanced capacity to manage the risks of deployed AI systems and to allocate risk management resources based on risk measures. **When serious consequences that impact life or liberty are detected, affected stakeholders should have a role in decisions to supersede, disengage, or deactivate AI systems when they are deployed on behalf of the government upon its constituents.** It is incumbent on Framework users to continue to apply the Manage function to deployed AI systems as methods, contexts, risks, and stakeholder expectations evolve over time."

**Facilitating Transparency in Criminal Investigations and Prosecutions**

Risk management of criminal investigations and criminal prosecutions requires special transparency considerations. In the criminal process, algorithm failures jeopardize a person's life

[18] Cynthia Lum, Christopher S. Koper & James Willis, *Understanding the Limits of Technology's Impact on Police Effectiveness*, 20 Police Q. 135 (2017); Malcolm Sparrow, *A different type of work. The Character of Harms*, *in* Operational Challenges in Control (2008); Eric L. Piza, Sarah P. Chu & Brandon C. Welsh, *Surveillance, Action Research, and Community Oversight Boards: A Proposed Model for Police Technology Research*, *in* The globalization of evidence-based policing. Innovations in bridging the research-practice divide (Eric L. Piza & Brandon C. Welsh eds., 2021).

[19] Barry Friedman, Unwarranted: Policing without Permission (2017).

and liberty. If these systems are going to be used by the government to investigate and prosecute people accused of crimes, there must be transparency; people facing criminal charges must be informed that AI or automated technologies have been used to accuse them of crimes, and have access to the algorithms and source code underlying these technologies. The accused must also be provided with an understanding of the processes and procedures of any human element involved in the technology and, of course, the research purporting to validate the technology. The AI RMF recognizes the "relationship between risk and accountability associated with AI and technological systems more broadly differs across cultural, legal, sectoral, and societal contexts" (p. 15). Given this understanding, the AI RMF should include language that specifically recognizes that when life and liberty are at stake, transparency must increase.

**Recommendations:**

- Add additional language in the second paragraph of Section 4.5 emphasizing that as consequences become more severe, transparency measures should also increase.

> "Determinations of accountability in the AI context relate to expectations of the responsible party in the event that a risky outcome is realized. The shared responsibility of all AI actors should be considered when seeking to hold actors accountable for the outcomes of AI systems. The relationship between risk and accountability associated with AI and technological systems more broadly differs across cultural, legal, sectoral, and societal contexts. **The more severe the consequence, such as when life and liberty are at stake, AI and automated technology developers and those who implement them must proportionally and proactively increase their transparency practices such that the design, deployment, activity, and outcomes are made visible to affected individuals.** Grounding organizational practices and governing structures for harm reduction, like risk management, can help lead to more accountable systems."

**Duty to Correct and Notify**

The AI RMF should more explicitly integrate the duty to correct and notify in how to respond to negative outcomes. The duty to correct and notify is an ethical and professional obligation of criminal legal system stakeholders when an adverse event occurs.[20] 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.

[20] Sarah P. Chu, *Duty to Correct and Notify*, *in* Encyclopedia of Forensic Sciences (Max Houck & Kevin Lothridge eds., Third ed. Forthcoming); Barry Scheck, *The Integrity of Our Convictions: Holding Stakeholders Accountable in an Era of Criminal Justice Reform*, 48 Georget. Law J. Annu. Rev. Crim. Proced. iii (2019).

For example, there is currently insufficient research to help us understand the consequences of the "human in the loop" in the use of facial recognition technologies and psychological factors like the limits of human memory and implicit bias. Should research in this area one day reveal severe disparities in the role of the human in the loop, we may need to review cases that fall into these risk categories. A similar review should be undertaken for new versions of code or changes to databases that influence the outcome of AI and automated technology analyses.

The duty to correct and notify is consistent with the AI RMF's recognition that "inscrutable AI systems can complicate the measurement of risk" (p. 9), "[t]ransparency is often necessary for actionable redress related to AI system outputs that are incorrect or otherwise lead to negative impacts" (p. 15), and "[s]trategies to maximize benefits and minimize negative impacts are planned, prepared, implemented, and documented, and informed by stakeholder input" (p. 25). The AI RMF can improve the attention to duty to correct and notify by explicitly referencing it in Section 6.4.

**Recommendations:**

- Previously we suggested additions to the second paragraph of Section 6.4. Manage (p. 25). Those suggested changes will remain in **bold**. Additional suggested language is added here in ***bold italic***:

> "After completing the Manage function, plans for prioritizing risk and continuous monitoring and improvement will be in place. Framework users will have enhanced capacity to manage the risks of deployed AI systems and to allocate risk management resources based on risk measures. **When serious consequences that impact life or liberty are detected, affected stakeholders should have a role in decisions to supersede, disengage, or deactivate AI systems when they are deployed on behalf of the government upon its constituents. *Public agencies should subsequently implement a transparent process to implement the duty to correct and notify.*** It is incumbent on Framework users to continue to apply the Manage function to deployed AI systems as methods, contexts, risks, and stakeholder expectations evolve over time."

**Incentives for Adoption**

NIST has the unique opportunity to provide incentives to AI and automated technology developers and implementers to adopt AI RMF. Although NIST is not a regulatory agency, it can motivate developers and implementers of AI RMF by elevating them or certifying entities that have demonstrated a robust implementation of the AI RMF.

**Conclusion**

Thank you in advance for your consideration of the feedback we respectfully offer. The Innocence Project continues to appreciate NIST's consistent transparency and efforts to engage stakeholders in their efforts to improve the trustworthiness of algorithmic technologies. We look forward to working with NIST to advance equity in science and technology in the criminal legal

system to ensure their simultaneous contributions to public safety, strengthening communities, and the just and equitable administration of justice.