Electronic Frontier Foundation Feedback
on the NIST AI Risk Management Framework and Playbook

Electronic Frontier Foundation is a San Francisco-based nonprofit that focuses on litigation, policy, and research to protect and advance civil rights at the intersection of digital and technical developments. We appreciate the opportunity to provide feedback on NIST's AI Risk Management Framework and Playbook and offer the following comments and suggestions for these materials as they exist in draft form.

General comments and suggestions

• **Don't undersell the real harms AI systems can cause. —** The AI RMF frames civil/human rights and economic risk equally. This is an inappropriate equivalence. While AI can contain flaws that can have negative economic consequences for those using it, the risks to human rights and human life that AI can pose in some contexts is important to emphasize. Appendix B describes these concerns, and the content there, particularly the final paragraph about the misperception that AI works, should be highlighted in the body of the report, somewhere near the very top.

• **Integrate the terms "machine learning" and "algorithm" earlier in the AI RMF.** — AI is a term that can oversell the complexity of a particular system, and some of the most dangerous uses of AI are those that would be better described as an algorithm[1]. It should be made clear that AI in this context can encompass systems referenced as an "algorithm" or "machine learning," preferably as part of the early section describing the use of "AI system" in the AI RMF.

• **Plan for outreach to AI actors. —** Government and private entities should consider many of the elements in this framework and their social responsibilities when it comes to developing, adopting, and utilizing AI. Educating these entities will require outreach. We encourage that effort and appreciate NIST's work on these materials. There are many places where those developing and selling these technologies gather, collecting a lot of other types of information; these concepts should be among them. There will also be individuals within organizations that will want to consider the AI RMF but may not be in a position to implement its concepts organizationally. These AI actors (including those in AI Design, AI Development, and AI Deployment) may benefit from resources in the Playbook that are tailored to their roles.

• **Clear information sharing on known flaws and biases is crucial.** — Responsible parties should communicate these known errors to other stakeholders and users. The AI RMF should emphasize this need for communication on these issues where possible.

• **Government agencies should make AI systems explainable to the public.** — Government use of AI should have a benchline expectation that officials can and will be able to describe the AI and data lifecycle to the public it serves.

• **Measures of effectiveness.** — AI users should devise goals for the system's implementation and utilize some metrics of effectiveness. These should be reviewed regularly and be made available to the public.

Elements missing from the framework and suggestions for addressing them.

• **Emphasize the need for a harm response strategy to take harmful systems out of commission.** — While some ideas of bias are culturally-dependent or otherwise variable, when an AI system or the data on which it is built has been found to violate standards of equity or civil rights, use of that system needs to stop. For example, some predictive policing systems have been built on policing data created from departments with known unfair policing practices. Such data should not be implemented into future AI systems, and the AI RMF should emphasize in the Govern 1 section the need for appropriate procedures to deal with these situations. (The language in Govern 6.2 speaks to this kind of need for a response when harmful AI or data has been identified.) We support the recommendation, on page 9, that "where negative impacts are imminent,

[1] https://apnews.com/article/child-welfare-algorithm-investigation-9497ee937e0053ad4144a86c68241ef1

severe harms are actually occurring, or catastrophic risks are present — development and deployment should cease in a safe manner until risks can be sufficiently mitigated." This sentiment should be emphasized in the RMF.

• **Highlight the government's responsibilities when AI use would constitute a new "substantive rule" and other legal requirements.** — Use of AI and other algorithmic systems is not like use of "other" tools and, particularly when used by government entities, not simply a procurement decision akin to buying a printer. AI systems are decision-making machines that can have serious impacts on the lives, wellbeing, and civil liberties of the people subject to their determinations. For this reason, it is incumbent upon government entities to consider whether use of an AI system or an algorithm represents the creation of a new substantive rule, which requires public notice and comment. There may be other legal obligations that are implicated by the adoption or use of an AI system. We recommend that the Playbook collect and highlight some of the legal considerations the government and other entities must consider in various contexts.

• **Underline the risks associated with data collection.** — AI systems built on flawed, biased, or otherwise dirty data lead to outcomes built on harmful precedents. We have seen, for example, law enforcement agencies use AI systems, like predictive policing, built on policing practices that violate civil liberties and face recognition systems built without a single person of color in their training sets. Injustices meted upon populations already vulnerable in society are exacerbated by such errors. Some version of Map 5.1 ("Potential positive and negative impacts to individuals, groups, communities, organizations, and society are regularly identified and documented") should appear in the Map 1 section ("Context is established and understood").

• **Mention the pitfalls of human factors.** — Humans-in-the-loop can contribute to or launder faulty AI findings. For example, a gunshot detection system might require the HITL to decide in a matter of seconds whether a sound is a gunshot and worth the deployment of police, creating opportunities for human error in an environment fraught with corporate pressure. The Playbook should include examples of the HITL in context and some known unsuccessful placements in the AI process.

• **Underscore the need for clear communication about data infiltration.** — Many AI systems used by government agencies and others are built on incredibly sensitive information — location, public services received, etc. — and security breaches of these systems need to be communicated to those whose information has been accessed and other entities responsible for protecting this data. This should be called out explicitly.

Other relevant EFF resources

We invite you to explore our other resources:

- EFF Issue Page — Artificial Intelligence & Machine Learning[2]
- The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation[3]
- Police Use of Artificial Intelligence: 2021 in Review[4]
- Artificial Intelligence and Algorithmic Tools: A Policy Guide for Judges and Judicial Officers[5]

[2] https://www.eff.org/issues/ai
[3] https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf
[4] https://www.eff.org/deeplinks/2021/12/police-use-artificial-intelligence-2021-review
[5] https://www.eff.org/files/2018/12/21/ai_policy_issues_handout.pdf