PARTNERSHIP ON AI

AI-Standards, ℅ Elham Tabassi
National Institute of Standards and Technology
100 Bureau Drive, Stop 200
Gaithersburg, MD 20899

The Partnership on AI Response to the National Institutes of Standards and Technology [Request for Information on Artificial Intelligence Standards](#)

Dear Ms. Tabassi:

The Partnership on AI (PAI) is a global multistakeholder organization that brings together academics, researchers, civil society organizations, companies building and using AI technology, and other groups working to realize the promise of artificial intelligence. The Partnership was established to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society. Today, PAI convenes more than 90 partner organizations from around the world to be a uniting force for the responsible development and fielding of AI technologies. The Partnership staff composed this response. This document should not be taken as stating the view of any particular member organization of the Partnership on AI.

PAI and its Partners create best practices for AI technologies in the areas of labor and the economy; the social and societal influences of AI; safety-critical AI; the collaborations between humans and AI systems; AI for social good; and fair, transparent, and accountable AI. Two recent reports exemplify PAI's work: a [Report on the Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System](#) and the [AI, Labor, and the Economy Case Study Compendium](#). PAI is also establishing common transparency standards among our Partner organizations, called "Annotation and Benchmarking on Understanding and Transparency of Machine learning Lifecycles," or [ABOUT ML](#). We are happy to provide additional information about the workshops,

convenings, and other activities we conduct to ensure AI benefits people and society.

One of the biggest challenges in developing standards for AI technologies is the rapid pace of innovation in this field and the wide scope of domains where they have been implemented. The deliberative nature of standards development can complicate organizational attempts to publish standards that reflect the current state of technological development. It is for this reason that we applaud NIST's current approach, namely to develop a plan for creating standards while continuing to participate in the International Standards Organization/International Electrotechnical Commission Joint Technical Commission 1/Sub Committee 42 (ISO/IEC JTC1/SC42) processes, and engaging the broader AI research community, including through multistakeholder organizations like the Partnership on AI, in its considerations and approach for standards development.

While compelling reasons exist to think carefully before embarking on standards development in dynamic, technically evolving fields, the time may be appropriate to focus these efforts on specific high-stakes applications of AI technologies. Today, AI systems are being implemented in courtrooms, police departments, loan applications, hiring decisions, housing applications, and many other sectors before they have been examined for harms to underrepresented, vulnerable, and marginalized populations.

One example is the deployment of risk assessment tools in the criminal justice system. Although the use of these tools was in many cases motivated by the desire to bring greater objectivity to the criminal justice system, researchers have since identified the substantial risk that these tools engage in flawed decision-making and entrench rather than alleviate existing societal biases. Moreover, without sufficient transparency around the tools and their design and technical training for judges, lawyers, and clerks working with these tools, they raise additional procedural fairness concerns, as judges and prosecutors

can easily misinterpret the predictions of these tools and defendants are likely ill-equipped to contest the algorithmic decisions.

PAI's [report on risk assessment tools](#) reflects the views and expertise of PAI's partner community and outlines ten minimum requirements that risk assessment tools should meet before deployment (see Appendix for the list of requirements). Because no existing risk assessment tool meets all ten requirements, the consensus view of our Partners was that current tools should not be used to automate pre-trial detention decisions.

The ten requirements fall into three categories: (i) technical challenges related to accuracy, validity, bias, (ii) human-computer interface issues that reflect the ways in which judges, clerks, and lawyers in the criminal justice system understand and use these tools, and (iii) governance, transparency, and accountability concerns that stem from the fact that these tools automate legal and policy decisions. These ten requirements are meant to serve as a roadmap of areas to address with standards rather than standards in and of themselves. Because of the complex ethical trade-offs involved with setting appropriate standards regarding individuals' life and liberty, precise standards can only be accomplished through collaboration between policymakers and experts.

Our report is thus an invitation for further engagement with policymakers and toolmakers to set standards to ensure that any continued or future use of risk assessment tools addresses the technical and ethical challenges that our Partners highlight. PAI serves to foster such collaboration across the research community, civil society organizations, developers of AI, and policymakers through our convenings and research projects.

Although PAI's report focused on risk assessment tools in the criminal justice system, many of the requirements outlined in that document are relevant to applications of AI to other domains. The concerns flagged in our report and the lessons learned from local jurisdictions' deployment of risk assessment

tools can thus inform broader conversations about government use of automated tools in high stakes settings. As NIST advises federal agencies on setting standards for AI tools, we urge caution, introspection, and further study of the impacts of sensitive application areas prior to public deployment and consideration of the principles outlined in our report. We look forward to future collaboration with NIST and other entities interested in promoting responsible development and use of AI in PAI work to come.

Regards,

Terah Lyons
Founding Executive Director

PARTNERSHIP ON AI

APPENDIX:

Minimum Requirements for the Responsible Deployment of Criminal Justice
Risk Assessment Tools
- Accuracy, Validity, and Bias
  - Requirement 1: Training datasets must measure the intended variables
  - Requirement 2: Bias in statistical models must be measured and mitigated
  - Requirement 3: Tools must not conflate multiple distinct predictions
- Human-Computer Interface
  - Requirement 4: Predictions and how they are made must be easily interpretable
  - Requirement 5: Tools should produce confidence estimates for their predictions
  - Requirement 6: Users of risk assessment tools must attend trainings on the nature and limitations of the tools
- Governance, Transparency, and Accountability
  - Requirement 7: Policymakers must ensure that public policy goals are appropriately reflected in these tools
  - Requirement 8: Tool designs, architectures, and training data must be open to research, review, and criticism
  - Requirement 9: Tools must support data retention and reproducibility to enable meaningful contestation and challenges
  - Requirement 10: Jurisdictions must take responsibility for the post-deployment evaluation, monitoring, and auditing of these tools