Palantir

July 19, 2019

National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899

To Whom It May Concern:

We thank the National Institute of Standards and Technology (NIST) for providing an opportunity to review and comment on the draft Plan for Federal Engagement in AI Standards.

Palantir Technologies Inc. ("Palantir Technologies," "we") is a software company that builds data integration and analysis platforms for critical institutions spanning the government, commercial, and non-profit sectors. While we are not fundamentally an artificial intelligence (AI) company, we view AI techniques as end points of our analytics platforms that can be applied to certain classes of problems to empower humans. Additionally, many of our partners use Palantir Foundry, our data management platform, as a lever for effectively and responsibly developing and operationalizing AI towards their core mission sets. Through our work and experience, we have developed a number of broad insights into the process of developing, training, and operationalizing AI that we believe may provide useful points of departure in refining the Plan for Federal Engagement in AI Standards.

In supporting customers who wish to build out their AI capabilities, we have seen up close how AI applications are configured, trained, and operationalized in a variety of different contexts. From this experience, we recognize that all phases of AI implementation — from algorithm design to deployment — are subject to both (1) inherent limitations (e.g., bad data, biased data, incomplete objective functions, incomplete codification of real world constraints) and (2) extrinsic failings of designers, developers and operators (e.g., bias, unintentional errors).

As such, AI techniques must be examined at all phases with respect to not only the veracity, viability, and efficacy of the AI technology itself, but also (1) the fidelity and trustworthiness of the data with which the AI are trained upon and (2) the purposes for which they are employed.

Given our belief that each phase of AI implementation should be critically examined, there are several critical, broad insights that we wish to present to NIST facilitators for consideration in inclusions on future drafts:

1. To enable AI success across all industries, establish general *principles* as a complement to vertical *standards*

The current draft's attempts to simultaneously optimize on **both** Horizontal (i.e., cross-sector) standards **and** Vertical (i.e., sector specific) standards strikes us as impractical and misguided. Our manifold experiences have underscored for us that insofar as AI is meant to augment skilled practices in various sectors, it must be mindful of the critical role of sector-specific subject matter expertise, histories, legal, regulatory, and other specialized considerations. All of these considerations have direct implication on the applicability of standards and suggest to us that standards will be most meaningful and practicable if they are established on a per-industry basis. General principles, on the other hand, may be reasonably established to cut across sectors and span most applications of AI. The distinction between principles and standards should be cleanly articulated, and that may imply a directional shift in parts of this document. We define standards as a framework that is, in theory, sufficient for guiding, directing, or resolving a particular issue. We define principles as more general, high-level rules or laws that provide guidance for the interpretation or application of a standard. Principles better lend themselves to broad or Horizontal use. Standards, by virtue of specificity, are most useful/meaningful in context or in Vertical use. We have submitted comments on the specific areas where believe the document may be modified to address this suggestion, including p. 6, lines 115 - 121 (see attached comments matrix).

2. Data management is essential to AI and should be included in any AI principles established

End-to-end data management plays a critical role in designing, building, and deploying trustworthy AI. Some data management features such as auditability and interpretability have been addressed in the supporting literature and are partially treated in the present draft (we also provide more detailed thoughts on the **human** component of auditability in the final point below). However, others are subtle and appear not to have been addressed in any form. In particular, we wish to highlight the importance of defining data management and further refining and developing the following data management points:

- **Data provenance:** All data involved in developing, training, and deploying AI should be tracked. This includes both upstream data (e.g., AI training inputs) and downstream data (e.g., outputs of systems that consume AI and/or combine AI outputs with other data). Including principles grounded in requiring comprehensive data provenance will direct the AI community to create more transparent, explainable AI.
- **Model Branching:** Models should be able to be branched or simulated so that you can ask "what-if" questions with different inputs/parameters/conditions, and understand how those inputs affect model outputs, and in turn, the decisions and analysis downstream of those outputs. Simulations should be able to run based on (a) changes in upstream data, (b) specific scenarios of interest, and (c) changes in downstream usage. Simulation is a more rigorous method of empirically assessing outcomes. This is a necessary supplement to AI transparency measures that are focused on internal process details.

3. AI models require human auditability as a component of rigorous evaluations and tracking

Often, AI systems are optimized as software systems rather than as data systems. This can be a function of design patterns (e.g., aggressive encapsulation and automation), as well as interfaces (e.g., data formats optimized for consumption by software rather than optimized for human analysis). While this enables software developers to collaborate and build complex systems together, it also raises the bar for what is needed in terms of making AI transparent. We advocate for making AI systems auditable by both software and humans, which means breaking down an end-to-end problem into steps to the extent possible, thereby creating the ability to rigorously interrogate those and any intermediate steps of a decision. An important point to draw out is that practical applications of AI models require the ability to methodically track and understand how input variations impact outputs. This is important to getting to the root of what is driving an AI model. For example, you will not be able to understand how a TV works by simply looking at individual transistors. Rather, you understand it by pressing buttons and seeing what each one does. Similarly, AI algorithms need to be testable against their outputs, with each permutation methodically tracked and documented.

It is our hope that the above high-level recommendations for this document will be of value in ensuring the success this effort. In addition to these broad suggestions, we have also submitted granular comments addressing specific sections of the document.

Thank you once more for the opportunity to contribute.

Sincerely,

Courtney Bowman
Privacy and Civil Liberties Lead
Palantir Technologies Inc.

Anirvan Mukherjee
Artificial Intelligence/Machine Learning Lead
Palantir Technologies Inc.