FINDINGS:
Towards Standards for Data Transparency for AI Models

The National Artificial Intelligence Advisory Committee (NAIAC)

April 2024

DRAFT

INTRODUCTION

Recognizing the significance of data in the development of trustworthy AI systems, the NAIAC working group on Trust, Safety, and Rights convened a public briefing on February 22, 2024 to explore the benefits, risks, and tradeoffs of creating and instituting baseline standards for data transparency for model creators.

Four expert speakers, listed below, answered the prompt: "What minimum standards for data transparency should there be for model creators?"

Dr. Meghan Dierks, Chief Data Officer at Komodo Health and Assistant Professor at Harvard Medical School

Jon Iwata, Founding Executive Director of the Data and Trust Alliance

Dr. Yacine Jernite, Machine Learning and Society lead at Hugging Face

Jeffery Smith, U.S. Department of Health and Human Services, Office of the National Coordinator for Health IT

The conversation highlighted several themes. These findings reflect the experiences and expertise of the speakers from this briefing.

- As a topic, the question of data standards and data transparency is important to understanding and regulating AI. Knowing what kind of data a model was trained on is pertinent to being able to deploy that model in a trustworthy manner

- There is a lack of shared definitions and standards for data transparency, but there exist examples to learn from in industry and healthcare

- There are tensions to keep in mind when designing transparency standards. However, when such tensions are managed, these standards can create meaningful business and organizational benefits — in addition to enhancing the trustworthiness of AI outcomes

- Such standards benefit from multi-stakeholder input and ongoing stewardship to be effective

DRAFT

FINDINGS

Finding 1:
As a topic, the question of data standards and data transparency is important to understanding and regulating AI. Knowing what kind of data a model was trained on is pertinent to being able to deploy that model in a trustworthy manner.

Dr. Yacine Jernite of Hugging Face commented that current regulatory discussions and general discourse often overlook the critical role of data in AI development, comparatively focusing attention more on AI models and applications. However, understanding the quality of datasets that train models is a key input into deploying them for trusted outcomes — and developing an understanding of the circumstances in which such models are likely to succeed or fail. As such, establishing minimum standards for data transparency for model developers, along with creating appropriate tools for dataset examination, would be an important component of creating a responsible AI ecosystem. Dr. Jernite added the perspective that current evaluation frameworks for models are not robust or mature enough to guarantee trustworthy outcomes from those models; as such, understanding the qualities of training data remains a key factor in making responsible decisions about model deployment.

Finding 2:
There is a lack of shared definitions and standards for data transparency, but there exist examples to learn from in industry and healthcare.

Currently, there are no consistent standards or universal definitions for data transparency for AI model creators. Through the briefing, a range of definitions and approaches as to what data transparency could mean were shared. One speaker defined minimum transparency as having the ability as a model developer to assess the completeness, appropriateness, and suitability of a dataset for a model's intended use. Another cited other characteristics of data as important to include in a standard, such as dataset sizes, purposes, licensing deals, and more.

While it is early for the topic of data transparency for AI, there is much to be learned from industry and existing standards that hold similar purposes. Jon Iwata of the Data and Trust Alliance shared their eight data provenance standards[1] which aim to surface a baseline set of contextual metadata applicable across industries to address

[1] "Data Provenance Standards," Data & Trust Alliance, 2023. https://dataandtrustalliance.org/our-initiatives/data-provenance-standards#:~:text=The%20eight%20proposed%20Data%20Provenance,metadata%20field%20has%20associated%20values.

DRAFT

the need for practical data transparency standards. These standards include visibility into the data's source, legal rights, privacy and protection, generation date, data type, generation method, intended use, and restrictions and lineage.

Another sector to look to is healthcare where data transparency, which is used to help with accuracy, testing, and traceability, has long been balanced with concerns like privacy and access to sensitive data[2]. Jeffrey Smith of the U.S. Department of Health and Human Services shared related global healthcare standards as examples, like the Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) Final Rule.[3] This standard established transparency requirements for algorithms in health IT aiming to "promote responsible AI and make it possible for clinical users to access a consistent, baseline set of information about the algorithms they use to support their decision making and to assess such algorithms for fairness, appropriateness, validity, effectiveness, and safety." It also "requires support for an 'internet-based method' for patients to request a restriction on the use or disclosure of their data."

Finding 3:
There are tensions to keep in mind when designing transparency standards. However, when such tensions are managed, these standards can create meaningful business and organizational benefits — in addition to enhancing the trustworthiness of AI outcomes.

When designing minimum data transparency standards for AI model creators, two key tensions were discussed: 1) privacy, particularly the need to safeguard sensitive demographic or personal data, and 2) protecting trade secrets or proprietary information. On both of these points, learning from domains such as healthcare may be instructive.

For example, Dr. Megan Dierks of Komodo Health highlighted the delicate balance between preserving enough context in the data for it to be useful and protecting individual privacy to prevent the risk of re-identification. This is something that has long been done in the healthcare industry and is an area that would benefit from further discussion and learning from both where things have gone well and where standards have failed to protect from re-identification.

---

[2] For more information about the challenging balance of data for civil rights protection, see also NAIAC's recommendation on Data Challenges and Privacy Protections for Safeguarding Civil Rights in Government

[3] "Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing," Health and Human Services Department, Federal Register, January 2024. https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and.

DRAFT

Likewise, Mr. Smith expressed that a key concern from both dominant and emergent players in formulating data transparency standards in the healthcare space was the risk of exposing trade secrets and proprietary information[4]. However, he relayed that it became clear over time that there is a way to create minimum standards for transparency that prevent against this risk. As Mr. Smith said, "We're not asking industry to give away the store, we're simply asking where the store is located." To that end, he highlighted the benefits of model cards, or "nutrition labels" for model transparency and understanding. Creating a consistent data standard would help determine the ingredients for these model cards.

Speakers also highlighted the benefits unlocked by successfully navigating these tensions and enabling consistent standards around minimum data transparency. For example, Mr. Iwata reminded briefing participants that businesses typically have teams that invest large amounts of time preparing data for use in their enterprise applications. The creation of standard approaches could save businesses considerable time and resources, and make the process of establishing a minimum standard of transparency easier.

Finding 4:
Such standards benefit from multi-stakeholder input and ongoing stewardship to be effective.

Finally, the briefing underscored the importance of ongoing input and stewardship from a wide range of stakeholders to develop data transparency standards that will benefit all parties involved. Dr. Dierks and Mr. Iwata stressed the importance of having the right individuals at the table — from multiple perspectives — for the development of standards. This ideally includes businesses and organizations who would be asked to adopt such standards and organizations representing communities that may be impacted by such standards. Dr. Jernite stressed the importance of ongoing stewardship and maintenance of such standards, including by a standards body that can review how the standard is performing over time and improve on that performance based on learnings.

DRAFT

ACKNOWLEDGEMENTS

The **NAIAC working group on xx** participated in the preparation of this document. Contributors include:

[4] An additional tension that could work against data transparency is the potential lack of existing clarity on the application of particular copyright standards around fair use to generative AI.

- **xx**

The full membership of NAIAC reviewed and approved this document.

ABOUT NAIAC

The National Artificial Intelligence Advisory Committee (NAIAC) advises the President and the White House National AI Initiative Office (NAIIO) on the intersection of AI and innovation, competition, societal issues, the economy, law, international relations, and other areas that can and will be impacted by AI in the near and long term. Their work guides the U.S. government in leveraging AI in a uniquely American way — one that prioritizes democratic values and civil liberties, while also increasing opportunity.

NAIAC was established in April 2022 by the William M. (Mac) Thornberry National Defense Authorization Act. It first convened in May 2022. It consists of leading experts in AI across a wide range of domains, from industry to academia to civil society.

https://www.ai.gov/naiac/

###

DRAFT