Comments of the World Privacy Forum regarding National Institute of Standards and Technology (NIST) Request for Information (RFI) Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence (Sections 4.1, 4.5, and 11), Docket Number: 231218-0309

Sent via email to ai-inquiries@nist.gov

ATTN: AI E.O. RFI Comments
National Institute of Standards and Technology
100 Bureau Drive Mail Stop 8900
Gaithersburg MD 20899–8900

The World Privacy Forum is pleased to submit comments regarding the National Institute of Standards and Technology (NIST) Request for Information (RFI) Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence (Sections 4.1, 4.5, and 11), 88 FR 88368, https://www.federalregister.gov/documents/2023/12/21/2023-28232/request-for-information-rfi-related-to-nists-assignments-under-sections-41-45-and-11-of-the

The World Privacy Forum is a non-partisan 501(c)(3) public interest research group focused on conducting research, analysis, and education in the area of privacy and complex data ecosystems and their governance, including in the areas of identity, AI, health, and others. WPF works extensively on privacy and data governance across multiple jurisdictions, including the U.S., India, Africa, Asia, the EU, and additional jurisdictions. For more than 20 years WPF has written in-depth, influential research regarding systemic data issues. These include medical identity theft, India's Aadhaar identity ecosystem —peer-reviewed work which was cited in the landmark Aadhaar Privacy Opinion of the Indian Supreme Court — The Scoring of America, an early and influential report on machine learning and consumer scores. Most recently, WPF published *Risky Analysis,* a report on AI Governance Tools that establishes the beginnings of an evaluative environment for these tools. WPF co-chairs the UN Statistics Data Governance and Legal Frameworks working group, and is co-chair of the WHO Research, Academia, and Technical Constituency. At the OECD, WPF researchers participate in the OECD.AI AI Expert Groups, among other activities. WPF participated in the core group of AI experts that collaborated to write the OECD Recommendation on Artificial Intelligence, now widely viewed as

the leading normative principles regarding AI. WPF research on complex data ecosystems governance has been presented at the National Academies of Science and the Royal Academies of Science. See our reports and other data at World Privacy Forum: https://www.worldprivacyforum.org.

We have responded to the RFI in three discreet areas: I. implementation of NIST responsibilities to help the AI community in the safe and trustworthy development and responsible use of AI, II. reducing the risk of synthetic content, and III. global standardization efforts. We discuss global standardization efforts in regards to A. national governments, B. the development context, and C. indigenous contexts.

I. Assist in the implementation of NIST responsibilities to help the AI community in the safe and trustworthy development and responsible use of AI.

Our comments in this section focus on a critically important aspect of implementation of safe and trustworthy development and responsible use of AI, which is AI governance tools. These tools operate in the background of AI systems, automating the tasks of checking AI systems for various elements of trustworthiness. These tools often operate less visibly, and thus far, they have escaped much notice. As a result, there are meaningful deficiencies regarding evaluative environments for testing and standards setting, as well as deficiencies relating to the development of policy guidance. Given the central importance of AI governance tools, these are deficiencies worth correcting.

AI systems should not be deployed without simultaneously evaluating the potential adverse impacts of such systems and mitigating their risks. Most of the world agrees about the need to take precautions against the threats posed by AI systems. Tools and techniques exist to evaluate and measure AI systems for their inclusiveness, fairness, explainability, privacy, safety and other trustworthiness issues. These tools and techniques – which WPF calls collectively AI governance tools – can improve such issues. While some AI governance tools provide reassurance to the public and to regulators, the tools too often lack meaningful oversight and quality assessments. Incomplete or ineffective AI governance tools can create a false sense of confidence, cause unintended problems, and general- ly undermine the promise of AI systems. This report addresses the need for improved AI governance tools.

It is the goal of WPF to help gather evidence that will assist in the building of a more reliable body of AI governance tools. We began this process in 2023 with the publication of *Risky Analysis: Assessing and Improving AI Governance Tools —An international review of AI Governance Tools and suggestions for pathways forward*.[1] This report

[1] Kate Kaye, Pam Dixon *Risky Analysis: Assessing and Improving AI Governance Tools —An international review of AI Governance Tools and suggestions for pathways forward* World Privacy Forum, 15 December 2023. https://www.worldprivacyforum.org/2023/12/new-report-risky-analysis-assessing-and-improving-ai-governance-tools/.

analyses, investigates, and appraises AI governance tools, including *practical guidance*, s*elf assessment questionnaires*, *process frameworks*, *technical frameworks*, *technical code*, and *software* disseminated in Africa, Asia, North America, Europe, South America, Australia and New Zealand. The report also analyzes existing policy frameworks, such as data governance and privacy, and how they integrate into the AI ecosystem. In addition to an extensive survey of AI governance tools, the research presents use cases discussing the contours of specific risks. The research and analysis for the report connects many layers of the AI ecosystem, including policy, standards, scholarly and technical literature, government regulations, and best practices.

Our work found that AI governance tools used in most regions of the world for measuring and reducing risks and negative impacts of AI could introduce novel, unintended problems or create a false sense of confidence unless accompanied by evaluation and measurement of those tools and their effectiveness and accuracy. WPF suggests pathways for creating a healthy AI governance tools environment, and offer suggestions for governments, multilateral organizations, and others creating or publishing AI governance tools. These suggestions include best practices taken from existing AI and other quality assessment standards and practices already in widespread use. Appropriate procedural and administrative controls include:

1) providing AI governance tool documentation and contextualization, review, audit, and other quality assurance procedures to prevent integration of inappropriate or ineffective methods in policy guidance;
2) identifying and preventing conflicts of interest; and
3) ensuring that capabilities and functionality of AI governance tools align with policy goals. If governments, multilateral institutions, and others working with or creating AI governance tools can incorporate lessons learned from other mature fields such as data governance and quality assessment, the result will establish a healthier body of AI governance tools, and over time, healthier and more trustworthy AI ecosystems.

The *Risky Analysis* report and its associated data set, which will be uploaded as an addendum to these comments, is intended to begin building much-needed evidence and procedures regarding how to implement trustworthy AI by analyzing AI governance tools and their functions. AI governance tools, when they function well, can assist the people, businesses, governments, and organizations implementing AI or researching AI to delve into various aspects of how AI models are functioning, and if they are performing in expected or intended ways. However, when AI governance tools do not function well, they can exacerbate existing problems with AI systems.

A. Building an Evidence Base Regarding AI Governance Tools

AI governance tools are important because they can map, measure, and manage complex AI governance challenges, particularly at the level of practical implementation.

The tools are intended to remove bias from AI systems,[2] or increase the explainability of AI systems, among other tasks. Seeking an orderly, automated way of solving complex problems in AI systems can create efficiencies. But those same efficiencies, if not well-understood and appropriately constrained, can themselves exacerbate existing problems in systems and in some cases create new ones. This is the case with AI governance tools, an important and nascent part of AI ecosystems which this report defines as:

> **AI Governance Tools:**
>
> Socio-technical tools for mapping, measuring, or managing AI systems and their risks in a manner that operationalizes or implements trustworthy AI.[3] [4]

An AI governance tool can be used to evaluate, score, audit, classify, or improve an AI system, its decision outputs, or the impacts of those outputs. These tools come in many forms. This report classifies AI governance tools in the following categories: *practical guidance*, *self assessment questionnaires*, *process frameworks*, *technical frameworks*, *technical code*, and *software*.

While AI governance tools offer the promise of improving the understanding of various aspects of AI systems or their implementations, not all AI governance tools accomplish the goals of mapping, measuring, or managing AI systems and their risks, which we posit are essential features of an effective AI governance tool. Further, given the lack of systematic guidance, procedures, or oversight for their context, use, and interpretation, AI governance tools can be utilized improperly or out of context, creating the potential for errors ranging from small to significant. For example, AI

---

[2] An *AI system* is defined in the NIST AI Risk Management Framework as: "An engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy (Adapted from: OECD Recommendation on AI:2019; ISO/IEC 22989:2022)." *See*: NIST AI RMF, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. The OECD has updated its definition of an AI System as of 2023. The new definition is: "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment." Definition available at: *OECD AI Principles Overview*, OECD, https://oecd.ai/en/ai-principles.

[3] The definition for *AI governance tools* was developed by the authors of this report at the World Privacy Forum. It is based on the research for this report, the scholarly literature, and consultation with a wide range of technical, standards, legal, and policy experts. This definition maps to the OECD AI Principles, the National Institutes of Standards and Technology Trustworthy and Responsible AI principles, and the general outlines of the EU AI Act.

[4] The definition for *AI governance tools* excludes statutes, regulations, and common law.

governance tools can be used in novel or "off-label"[5] ways, which can lead to meaningful errors in contextualization and interpretation. Some of the more complex AI governance tools can create additional risk by producing a rating or score that in and of itself can be subject to error or misinterpretation, especially if there is a lack of documentation and guidance for use of the tool. All told, flawed usage and interpretation can result in a gap between what people want these tools to accomplish, and what these tools actually do accomplish.

WPF built an initial analysis of AI governance tools by conducting an extensive survey of the tools across multiple modalities and jurisdictions. We utilized the evidence from the survey of tools in conjunction with in-depth case studies and scholarly literature review to construct an extensive index of AI governance tool types. (The complete survey of tools is located in the full report.)

Some findings from the work include:

- **AI governance tools are already widely available and in use:** AI governance tools exist across Africa, Asia, Europe, North America, South America, and Oceania (Australia and New Zealand), at varying levels of maturity and dispersion. Governments, multilateral organizations, academia, civil society, business, and others utilize these tools in different types of AI implementations.[6] This research focuses on

---

[5] The term "off label use" originally stemmed from the practice in clinical settings of ing prescription drugs in a way that differs from what is approved by the FDA and printed on the original prescription label. In the AI context, "off-label" refers to the practice of taking a technology that was created for one context, and using it in another outside of the original use case. NIST mentions "off label use" in its AI Risk Management Framework: "…existing frameworks and guidance are unable to….consider risks associated with third-party AI technologies, transfer learning, and off-label use where AI systems may be trained for decision-making outside an organization's security controls or trained in one domain and then "fine-tuned" for another." *NIST AI Risk Management Framework,* National Institute of Standards and Technology, Feb. 2023. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. p. 39. In a study of off-label use of imaging databases, a National Academy of Sciences study found that the practice could lead to bias in AI algorithms. See: Efrat Shimron, Jonathan I. Tamir, Ke Wang, and Michael Lustig, *Implicit data crimes: Machine learning bias arising from misuse of public data*, March 21, 2022. PNAS, https://doi.org/10.1073/pnas.2117203119. And finally, increased risk is also associated with the term as used in its clinical context. See: Rebecca Dresser and Joel Trader, *Off-label prescribing: A call for heightened professional and governmental oversight*, Journal of Law and Medical Ethics, 2009 Fall: 37(3) 476-396. doi: 10.1111/j.1748-720X.2009.00408.x. "The potential for harm is greatest when an off-label use lacks a solid evidentiary basis. A 2006 study examining prescribing practices for 169 commonly prescribed drugs found high rates of off-label use with little or no scientific support."

[6] The findings are based on recent analysis of select tools. It is not the universe of all tools. All of the AI governance tools analyzed for this report address algorithmic fairness, discrimination and bias, and all but one addresses explainable, transparent and interpretable AI systems. Many of the remaining related items reviewed in Part II also address these two issues, which are prominent in AI governance.

AI governance tools used, promoted, or cataloged primarily by governments and multilateral institutions, especially those tools that seek to implement principles of trustworthy AI.[7] It remains difficult to quantify precisely how many tools exist.

- **Some AI governance tools feature off-label, unsuitable, or out-of-context uses of measurement methods:** More than 38% of AI governance tools reviewed in this report either mention, recommend, or incorporate at least one of three measures shown in scholarly literature to be problematic. These include off-label, unsuitable, or out-of-context applications when used to measure AI systems.[8]

- **Standards and guidance for quality assessment and assurance of AI governance tools do not appear to be consistent across the AI ecosystem: I**t became apparent during the research process that while some AI governance tool makers and toolkit providers and publishers have conducted some quality assessments of those tools, some have not; if they do conduct quality assessments, AI governance tool providers do not always conduct them according to an internationally recognized standard. Complete product labeling, documentation, provision for user feedback, requirements for testing, or provision of redress in the case of problems are important features of traditional products, but these features are not always present in AI governance tools.

B. The importance of Conducting Foundational Testing Work on AI Governance Tools

There is not enough data yet about how AI governance tools interface with specific standards. As a result, foundational work needs to be done to build an evaluative AI governance tools environment that facilitates validation, transparency, and other measurements. We urge NIST to undertake this work. Establishing an evaluation environment for AI governance tools will be crucial to create a healthy AI governance tools ecosystem, and more broadly, a healthier AI ecosystem.

In considering what might help build a transparent, evaluative environment for AI governance tools, the application of international and other standards holds potential.

---

[7] This research did not examine all tools available from academia or industry. By "principles of trustworthy AI," this research refers to, for example, the OECD Recommendation on AI and UNESCO Recommendation on the Ethics of Artificial Intelligence, UNESCO, adopted by 193 member states in 2021.

[8] Of the select 18 AI governance tools reviewed in detail in this report, 7—or more than 38%—mention or recommend using one of three problematic measures: fairness tools incorporating the US Four-Fifths or 80% Rule, or SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) for AI explainability. Each of these measurement methods have been shown to be unsuitable including when used in an "off-label" manner if applied to measure many types AI systems. See Part I for use cases describing these measures. *See also* Appendix C for a detailed accounting of this finding.

For example, the extensive quality assurance ecosystem articulated in formal standards and norms is well-understood across many mature sectors.

Although many established standards already exist and are important to acknowledge, currently, there is limited knowledge about the functionality of these standards as applied to AI governance tools. Testing of available tools would improve understanding of how existing standards might apply, and it would also support the ecosystem based on evidence. The Plan-Do-Check (or Study)-Act cycle will be a key tool to assist in this maturation.

C. Establishing Baseline Requirements for Documentation and Labeling of AI Governance Tools:

The research found high variability in the documentation and labeling of AI governance tools. This suggests that developing norms and standards regarding documentation and labeling of AI governance tools could produce meaningful levels of improvements. For example, it would be helpful if tools routinely include information about the developer, date of release, results of any validation or quality assurance testing, and instructions on the contexts in which the methods should or should not be used.

A privacy and data policy is also important and should be included in the documentation of AI governance tools. We request that NIST also undertake work on this issue so that there are commonly understood best practices for labeling and documentation of AI governance tools.

D. Key AI Governance Tools Use Case — SHAP and LIME: Popular but Faulty AI Explainability Metrics

(*Editorial note: In the full Risky Analysis report, WPF presents multiple use cases regarding certain recurring problems in some types of AI governance tools. Here, for these comments we present only the use case regarding the use of SHAP and LIME in AI governance tools due to space considerations*.)

In the absence of widely-adopted AI explainability standards, two approaches—SHAP and LIME—have grown in popularity, despite attracting an abundance of criticism from scholars who have found them to be unreliable methods of explaining many types of complex AI systems.[9]

---

[9] Dylan Slack et al., *Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods, AIES '20 Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Ass'n for Computing Machinery (Feb. 7, 2020), https://doi.org/10.1145/3375627.3375830.

Use of both SHAP[10] and LIME[11] has increased in part because they are model agnostic, meaning they can be applied to any type of model that data scientists build. An abundance of accessible and easy-to-use documentation related to the two methods has also fostered interest in them.[12]

The proliferation and adoption of SHAP and LIME as AI explainability methods recognized and used around the world is evident in documentation related to AI governance tools reviewed in Part II of this report. The review found that six AI governance tools from national governments reference or mention SHAP or LIME or both. In addition, a catalog of tools from a multilateral organization includes 12 items recommending SHAP and/or LIME.[13]

However, the applicability and efficacy of both SHAP and LIME are limited, particularly when used in an attempt to explain complex AI systems comprised of non-linear machine or deep learning models. In a typical use case, an AI practitioner might employ SHAP or LIME to explain a single instance of a model input, such as one decision or prediction, rather than the whole model. Because both methods work by approximating more complex, non-linear models (the types that are often called "black-box" models) with more straightforward linear models, they may produce misleading results.14

---

[10] *shap*, GitHub, https://github.com/shap.

[11] *lime*, GitHub, marcotcr, https://github.com/marcotcr/lime.

[12] This is based on a description of how SHAP and LIME work and their problems, as intended for a layperson, provided by Tim Miller, professor in artificial intelligence at the School of Electrical Engineering and Computer Science at The University of Queensland, during interviews conducted by WPF in June and November 2023. Miller was professor in the School of Computing and Information Systems at The University of Melbourne, and co-director of its Centre of AI and Digital Ethics, when WPF spoke with him in June 2023. In general, Miller said LIME is unstable and inappropriate as an explainability metric for machine learning, while SHAP-based methods are also limited in effectiveness. *Professor Tim Miller,* Univ. Of Queensland Australia, https://eecs.uq.edu.au/profile/9477/tim-miller.

[13] As noted in the findings of this report, several AI governance tools from national governments and multilaterals mention or recommend LIME and/or SHAP, including Chile's procurement form and process for government acquisition of algorithmic systems, IDB FairLAC's *Responsible use of AI for public policy data science handbook*, India's Responsible AI #AIFORALL Approach Document for India Part 1 – *Principles for Responsible AI*, Monetary Authority of Singapore's *FEAT Fairness Principles Assessment Methodology*, 12 items featured in the OECD's *Catalogue of Tools and Metrics*, and Singapore's *AI Verify*.

[14] November 2023 WPF interview with Tim Miller.

Short for Shapley Additive exPlanations, SHAP is based on a concept known as the the Shapley Value, introduced by Lloyd Shapley in 1951[15] in the context of cooperative game theory. The Shapley Value is a method used to determine the importance or contribution of each player to an overall competition between groups.16

Today, SHAP is used for another purpose entirely: in an attempt to expose and quantify feature importance, or the importance of factors that contribute to predictions of machine learning models.17 Oftentimes, SHAP is used in the hopes of revealing how factors affect the outputs of opaque, "black box" AI systems such as deep learning models and neural networks, which are difficult to interpret.

SHAP has grown in popularity since around 2017.[18] By 2020, use of SHAP for AI explainability had become widely adopted. When researchers asked people from 30 organizations in 2020 which explainability techniques they used and how, they reported that "feature importance was the most common explainability technique, and Shapley values were the most common type of feature importance explanation."[19]

1. Why SHAP and LIME Can Produce Misleading Explanations

SHAP reflects feature importance numerically. For instance, when using SHAP to determine how certain input features affect a more straightforward linear regression model trained on a California housing dataset, the SHAP value of the median house age in a block group might be expressed as -0.22, and the SHAP value of median income as +0.92. The process would be used to add other features, such as the

---

[15] Lloyd S. Shapley*, Notes on the N-Person Game — II: The Value of an N-Person Game*, RAND Corp. (1951), https://www.rand.org/pubs/research_memoranda/RM0670.html.

[16] S. Hart, *Shapley Value*, in The New Palgrave Dictionary of Economics 1-6 (1987), https://doi.org/10.1057/978-1-349-95121-5_1369-1.

[17] This description is based on an overview of how SHAPley Values work intended for a layperson as provided by Elizabeth Kumar, a Computer Science PhD candidate at Brown University, during interviews conducted by WPF in April and November 2023.
Lizzie Kumar personal website, https://iekumar.com/.

[18] Scott M. Lundberg & Su-In Lee, *A unified approach to interpreting model predictions, in Proceedings of the 31st International Conference on Neural Information Processing Systems*, Arxiv, 4768-4777 (Nov. 25, 2017), https://arxiv.org/abs/1705.07874 (a research paper presented at the NeurIPS conference in 2017 that is considered instrumental in popularizing the use of SHAP in AI explanations).

[19] Umang Bhatt et al., *Explainable machine learning in deployment, FAT* '20 Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency)*, Ass'n for Computing Machinery, 648–657 (Jan. 27, 2020), https://doi.org/10.1145/3351095.3375624.

average number of rooms or average home occupancy, until the current model output is reached.[20]

Although Shapley values have been applied in the context of feature importance for decades,[21] researchers have found several mathematical, practical, contextual, and epistemological problems associated with use of the method for explaining AI systems. For example, when attempting to attribute influence to a large set of features affecting AI model decisions or predictions, the approach relies on the modeler to decide which features count as "players" and which are redundant; these subjective decisions can affect the resulting explanations.[22]

Scholarly research also indicates that some users of SHAP may not understand how to interpret its results. A survey of data scientists using SHAP-based tools showed that many were unable to accurately describe what SHAP values or scores represented.[23] The study also found that the popularity of SHAP-based tools influenced some data scientists to trust the tools even if they did not understand what they did or how to interpret their results.

In addition, research shows that use of SHAP in AI explainability tools may lead users to falsely believe they discovered a precise explanation for why or how a system produced a specific output, such as a decision or prediction. This in turn may lead to misconceptions about what SHAP values represent and the actionable information that can be gleaned from them.[24]

---

[20] Vinícius Trevisan, *Towards Data Science*, Medium. Jan 17, 2022. https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137.

[21] W. Kruskal, *Relative importance by averaging over orderings*, The American Statistician, 41(1):6–10, 1987.

[22] I. Elizabeth Kumar et al., *Problems with Shapley-value-based explanations as feature importance measures.*

[23] Harmanpreet Kaur et al., *Interpreting interpretability: Understanding data scientists use of interpretability tools for machine learning, CHI '20 Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Ass'n for Computing Machinery, 114 (Apr. 23, 2020), https://doi. org/10.1145/3313831.3376219.

[24] Elizabeth Kumar et al., *Shapley Residuals: Quantifying the limits of the Shapley value for explanations*, Neural Info. Processing Sys. (2021).

Even scholars who acknowledge benefits of using SHAP to provide insight into certain aspects of models and data suggest they "can lead to wrong conclusions if applied incorrectly,"[25]and argue that they can be expensive to compute.[26]

LIME, a similar AI explainability method that has grown in adoption, was first introduced in 2016.[27]Short for Local Interpretable Model-agnostic Explanations, LIME produces explanations by randomly sampling "locally" around the singular instance chosen to be explained. But its randomness is a pitfall: If LIME is used again in an attempt to explain the very same instance, its explanation will be different.28 The use of LIME for AI explainability has been criticized, and research shows the method can lead to inaccurate results,29 or be manipulated or "gamed."[30]

Overall, the research indicating that there are vulnerabilities in these popular explainability measures is not reassuring; however, it is not completely unexpected. Trustworthy AI implementation is still nascent, with much work and refinement yet to come.

There are hints of further issues regarding SHAP, for example, some commonly used MLOps tools utilize SHAP.

E. Pathways for Building an Evaluation Environment and Creating Improvements in the AI Governance Tools Ecosystem

One of the most significant limitations of AI governance tools is the lack of knowledge about which contexts are and are not appropriate for the use of a particular tool. Further, even when some may be aware of the limitations of a tool, others using it may not be aware of the problems. To cite a specific example from the research, challenges in using SHAP for AI explainability are openly discussed amongst technical experts and

---

[25] Christoph Molnar et al., *General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models*, Arxiv (2022), https://arxiv.org/pdf/2007.04131.pdf.

[26] Christoph Molnar, *SHAP Is Not All You Need,* Mindful Modeler (Feb. 7, 2023), https://mindfulmodeler.substack.com/p/shap-is-not-all-you-need.

[27] Marco Tulio Ribeiro et al., *"Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Ass'n for Computing Machinery, 1135–1144 (Aug. 13, 2016), https://doi.org/10.1145/2939672.2939778.

[28] According to a November 2023 interview with Tim Miller.

[29] Romaric Gaudel et al., *s-LIME: Reconciling Locality and Fidelity in Linear Explanations*, Arxiv, (Aug. 2, 2022), https://arxiv.org/abs/2208.01510.

[30] Dylan Slack et al.

specialized researchers, however, this knowledge is not widely dispersed or understood. The research indicates several reasons why this and other breakdowns in contextual understanding, among other problems, are occurring. For example:

- AI governance tools are nascent; as such, a transparent, evaluative community basing their judgments on the evidence has yet not been fully constructed.

- The scrutiny and detailed research found in the scholarly literature has not reached all AI governance tool end users, tool publishers, or regulators.

- Some problems may be deeply encoded into the AI governance tools, and these problems can be very difficult to see by even careful researchers, much less by end users of the tools.

Recommendations:

1. We urge NIST to provide and help build testing environments for the evaluation of AI governance tools that will advance the measurement and practice of trustworthy AI. The research indicates that this is the most important foundational step that could be taken to enhance and ensure the quality of automated AI governance tools.
2. We request that NIST facilitate the development of consensus - based standards and best practices for AI governance tools based on the evidence, utilizing ethical guidelines and codes of conduct common to Standards Development Organizations and typical NIST standards development processes.
3. We urge NIST to develop guidelines for evaluation of AI governance tools throughout their use in the AI lifecycle.

II. Reducing the Risk of Synthetic Content

WPF understands the need for urgent work on solutions regarding content provenance and synthetic data. However, we are concerned about a race to solutions without an adequate evaluative environment that fully documents, tests, and explores the reliability of such techniques. We are also concerned about solutions that are proposed without concomitant study and documentation of the potential negative downstream impacts of such techniques. We have additional concerns about solutions that are being designated as quasi-standards, yet have not been built utilizing the normative ethical principles of conduct for Standards Development Organizations such as those NIST, ISO, IEEE, and ANSI, among other similar SDOs use.[31]

[31] ISO Code of Ethics and Conduct, as approved under Council Resolution 11/2023, adopted on 23 February 2023, ISO. https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100011.pdf .

Currently, WPF does not know of a reliable, standalone watermarking technique that is without flaws, as documented in the emerging scholarly literature on this topic.[32] WPF is interested in NIST effectuating an evaluative environment free of conflicts of interest where proposed techniques are fully studied, tested, documented, and discussed. It is based on this sound and robust empirical basis that any standard and best practices should be developed.

As mentioned, some techniques carry with them undesirable downstream consequences, including negative impacts on privacy as well as other problems. Regarding privacy, in addition to producing and embedding traceable data in content that shows who, where and when content and data was created, content and data provenance systems in development today could include unique identifiers and other identifiable information in an effort to validate, authenticate, certify or track the identity of content creators.[33]

Content provenance systems create the potential for every piece of digital text or social media post, artistic image, music, video or photo file to automatically carry with it an embedded, identifiable data trail.[34] For instance, camera software that automatically embeds information about the origin or provenance of a photo inside the digital photo file itself could include sensitive information such as the name of the camera owner or photographer, the date and time the creator took the photo, and the map coordinates of the location where the creator took the photo. Metadata added to the same file subsequently could include information identifying additional people in conjunction with time and location data by indicating who edited or changed the file, when, where and how. That identity related metadata could, depending on the system, include biometric data such as fingerprint, facial, and/or iris scans.

---

[32]Zhengyuan Jiang, Jinghuai Zhang, Neil Zhenqiang Gong, *Evading Watermark based Detection of AI-Generated Content,* In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS '23), November 26–30, 2023, Copenhagen, Denmark. ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3576915.3623189 . *See also* Github, zhengyuan-jiang / WEvade, public. https://github.com/zhengyuan-jiang/WEvade . See also: Melissa Heikkilä, *Why detecting AI-generated text is so difficult (and what to do about it)*, MIT Technology Review, February 7, 2023. https://www.technologyreview.com/2023/02/07/1067928/why-detecting-ai-generated-text-is-so-difficult-and-what-to-do-about-it/ .

[33] See for example documentation on JPEG Fake Media in he standards literature regarding JPEG. *See*: ISO/IEC JTC 1/SC29/WG1 N100388, REQ *Updated report on the JPEG Fake Media Call for Proposals,* 98th Meeting, Sydney, Australia, January 2023. https://jpeg.org/jpegfakemedia/documentation.html . "Standard assertions, along with optional use of W3C Verifiable Credentials, provide for specifying the identity of any/all actors along with the actions they performed, at what time and their location."

[34] Xiang, Ziyue and Horvath, Janos and Baireddy, Sriram and Bestagini, Paolo and Tubaro, Stefano and Delp, Edward J, *Forensic Analysis of Video Files Using Metadata*, in Proceedings of Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 10.1109/cvprw53098.2021.00115 .

Further, there are sectoral issues that have not been well-addressed yet. For instance, some systems in development today automatically apply metadata regardless of whether the data is protected through privacy-enhancing techniques and technologies, such as are frequently used with regulated health data. WPF urges NIST to set forth robust processes to evaluate these issues. Processes for assigning privacy labels or indicators in provenance metadata is an area of work that needs to be conducted with input from all stakeholders, including users, civil society and other groups affected by content provenance systems. The ethical principles of non-dominance discussed in these comments regarding Standards Development Organizations must be applied carefully to synthetic data and content provenance work.

WPF notes that NIST did not include retrieval techniques in the RFI. WPF supports this omission, and we hope it was intentional. In our early analysis, retrieval techniques introduce meaningful privacy challenges that could pose significant long-term problems and complications for data governance and data protection in AI ecosystems. For example, Jiang et al note at 5.4 that retrieval techniques can be accomplished on a Macbook Pro, which hints that the technique could be utilized readily at significant scale. The authors note that the use of retrieval as a technique poses " …potential privacy risk of exposing *all* LLM responses behind a binary classifier. [35] [emphasis by paper authors.]

There are additional downstream potentials for negative impacts. Deepfake detection methods vary in accuracy and may be built with imbalanced data of different races and genders [36] [37] that can result in large disparities in predictive performances across races. [38] Some researchers have begun to address these problems by proposing methods for improving fairness [39] and robustness [40] of existing deepfake detectors. It will be

---

[35] Z Kalpesh Krishna, Yixiao Song, Marzena Karpinska, John Wieting, Mohit Iyyer, *Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense*, 37th Conference on Neural Information Processing Systems (NeurIPS) 2023. https://arxiv.org/pdf/2303.13408.pdf .

[36] M. Masood, M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, and H. Malik, *Deepfakes generation and detection: State-of- the-art, open challenges, countermeasures, and way forward, Applied Intelligence*, pp. 1–53, 2022.

[37] Y.Xu, P. Terho¨rst, K.Raja, and M.Pedersen, *A comprehensive analysis of AI biases in deepfake detection with massively annotated databases*, arXiv preprint *arXiv:2208.05845*, 2022.

[38] Trinh, L., & Liu, Y., *An Examination of Fairness of AI Models for Deepfake Detection*. International Joint Conference on Artificial Intelligence, 2021. https://arxiv.org/abs/2105.00558

[39] Ju, Yan & Hu, Shu & Jia, Shan & Chen, George & Lyu, Siwei, *Improving Fairness in Deepfake Detection,* 2023 Arxiv, https://arxiv.org/pdf/2306.16635.pdf

[40] Nadimpalli, A.V., & Rattani, A., *On Improving Cross-dataset Generalization of Deepfake Detectors*. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 91-99. https://arxiv.org/abs/2204.04285.

important to ensure that detection methods do not produce unintended outcomes such as disparate negative impact on specific communities or groups of people. We note that proliferation of or requirements for registry with content provenance systems could compromise or stifle the creation and dissemination of content such as religious texts, controversial artwork, or journalistic reports from people fighting government repression. Inclusion of identifiable data in such systems could result in real-world dangers for victims of violence, marginalization of artists, thinkers and journalists, and self-imposed limits on free expression, including among children who may not pass content creator age requirements.

In addition, identifiable provenance data could travel along with content as it moves across borders, complicating approaches to trusted data flows.

In PAI's PAI's Responsible Practices for Synthetic Media, Section 2 recommends a number of principles for the developers of the technologies and infrastructures that support synthetic media and watermarking, including principle 4, which states:

> "Be transparent to users about tools and technologies' capabilities, functionality, limitations, and the potential risks of synthetic media."[41]

We agree, and would add to this transparency regarding the privacy and other potential risks of utilizing tools that detect synthetic media. Both sides of the problem are important to address, and this will be an important task for NIST.

And finally, we note that some of the standards development processes in the area of synthetic data and content provenance have exhibited the characteristics of a race to a solution, however, the race has not been inclusive of all stakeholders, and has in fact not always adhered to standards such as the ISO Code of Ethics and Conduct Principle 4, that is, "Promote and enable all voices to be heard (ex., participation in ISO activities is properly representative of the global community and that any barriers to full and equal participation are acknowledged and continuously and active addressed …) ". Further, some of the work has not complied with Principle 6: "Declare actual and potential conflicts of interest and manage them appropriately."

Any and all standards development in content provenance and synthetic media detection must occur within the ethical guidelines and codes of conduct common to all respected Standards Development Organizations. Standards developed in a non-dominant manner should not become the stem of work for a standard, and should not be adopted as a standard.

---

[41] *PAI's Responsible Practices for Synthetic Media, A framework for collective action*, Parntership on AI, 27 February 2-23. https://partnershiponai.org/wp-content/uploads/2023/02/PAI_synthetic_media_framework.pdf.

Recommendations:

1. An evaluative environment needs to be created to test for privacy and additional ethical considerations in content provenance and synthetic data detection systems.
2. Cross-border data flows, trust, and impacts on free speech are areas of policy impacts that need to be studied in this area of work.
3. Any content provenance or synthetic media standards that are utilized must comply with ISO, IEEE, ANSI, NIST and other respected SDOs Codes of Conduct and Principles. If there are standards that comply with OMB Circular A-119, these would also be acceptable given the parity of these requirements with such codes. However, informal "standards" build outside of ISO, IEEE, ANSI, and NIST standards of ethical requirements for standards including regarding conflicts of interest must not be used as the stems or early basis for formal standards making as these can introduce dominance and conflicts of interest, among other issues for which reasons the Codes of Conduct common to SDOs were created.

III. Advance Responsible Global Technical Standards for AI Development

There are meaningful challenges and opportunities regarding developing AI standards in the global context, including in the development context and the indigenous context. This section of comments distills initial thoughts regarding how NIST can begin to develop inclusive and respectful approaches to developing AI standards in these global contexts.

A. Global Standardization Efforts in AI regarding National Governments

ISO and NIST, among other standards bodies such as IEEE, have been deeply engaged in developing new standards in a variety of AI issue areas.[42] WPF acknowledges these efforts, and also acknowledges the importance of the ethical principles that guide this work.

As NIST knows, the work of ISO as well as NIST and other Standards Development Organizations (SDOs) follows a set of well-understood guidance, articulated in the ISO Code of Ethics and Conduct [43] with similar articulations at other SDOs. We provide an excerpt of the headlines of the ISO Code here: (*Note: Material in the parenthesis has been paraphrased to give an idea of the additional text for each principle, while remaining brief.*)

---

[42] For example, see *ISO Artificial Intelligence, top standards*, ISO. https://www.iso.org/sectors/it-technologies/ai . *See also* ISO/IEC 42001:2023, *Information Technology, Artificial intelligence Management system*. ISO. https://www.iso.org/standard/81230.html.

[43] ISO Code of Ethics and Conduct, as approved under Council Resolution 11/2023, adopted on 23 February 2023, ISO. https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100011.pdf .

1. Comply with legal and statutory obligations (ex, respect applicable laws and regulations and avoid collusive or anticompetitive behavior)
2. Perform and act in good faith, consistent with the purpose, policies and principles of the organization (ex., working for the benefit of the global community)
3. Behave ethically (ex., act with honesty, integrity, respect, openness and transparency in all dealings)
4. Promote and enable all voices to be heard (ex., participation in ISO activities is properly representative of the global community and that any barriers to full and equal participation are acknowledged and continuously and active addressed …)
5. Engage constructively in ISO activities
6. Declare actual and potential conflicts of interest and manage them appropriately
7. Protect confidential information
8. Protect ISO assets
9. Avoid and prevent any form of bribery or corruption
10. Escalate and resolve disputes and uphold agreed resolution

WPF agrees with these principles, and sees the need to ensure that any work to be conducted that is led by the U.S. follows these principles, with adaptations of principle 8 to be applicable to NIST instead of ISO. WPF would welcome a guidance note from NIST investigating implementation of these principles specific to global AI standards development prior to work on any global standards themselves.

We note that UNESCO, OECD, and NIST have already been collaborating on a variety of important AI projects. We would be pleased to see these collaborative efforts continue in the area of standards, ensuring that the work is spread across national governments in an equitable manner, ensuring that there is a balanced representation of jurisdictions and regions.

National Statistical Organizations (NSOs) have been engaged in the process of working in whole-of-government processes regarding emerging data and some AI issues. A series of UN background papers articulates the outcomes of a global task force that has been studying these issues and making recommendations.[44] Some of these recommendations could be useful for NISTs efforts in working with other governments regarding AI standards development.

Recommendation:

1. It would be highly beneficial for NIST to develop, in cooperation with multilateral organizations and national governments, standards and guidelines for developing global AI standards in an inclusive and respectful manner, following established

---

[44] *Data Governance: WS1, Statistics Poland and World Privacy Forum, Preliminary results from the Working Group on Data Stewardship*, United Nations Statistical Commission Fifty-third session, 1-4 March 2022, Item 3 of the provisional agenda. Items for discussion and decision: Data Stewardship, Pages 5-18. See also Equity and Inclusion, 18-23. https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3b-DSWG-E.pdf.

ethical best practices, with attention to regional, economic, social, technical, cultural, and other contextual factors.

B. Global Standardization Efforts in AI and the Development Context

In countries that are still developing capacity, work to ensure that best practices are used concurrently with the framework of SDGs to nurture progress in AI standardization efforts is key. However, much more work is needed to determine which best practices will be most effective in each context, and work that specifically identifies best practices that align with the SDGs as well as national data governance laws is still needed.

A multi-year project by the Center for Global Development and the World Privacy Forum found that too often, global standardization efforts led by the developed world resulted in a division of "standard makers" and "standards takers." [45] The reasons for this are complex, and a great deal of work has gone into understanding what might solve some of the inequities. This work which is ongoing.[46] Now that the "AI era" is in varied levels of development and implementation, the specific components of AI and what has often been called "big data" will need additional attention.

In the UN publication *A World that Counts*, the authors presciently included a discussion of minimizing the risks and maximizing the opportunities of the "data revolution." [47] The authors urged practitioners to determine methods of using data while safeguarding human rights, also mentioning algorithmic inferences. This was pioneering work. For standards work in the development context, there will need to be attention to the recommendations in *A World that Counts*, which we include by reference here. There will also need to be attention to the data governance laws present in each jurisdiction and region, which have crucially important interactions with AI, and again, attention to ongoing work regarding the SDGs and local and regional standards work.

---

[45] Michael Pisa, Pam Dixon, Ugonma Nwankwo, *Creating a level playing field for data protection*, Chapter, Development Cooperation Report, Shaping a Just Digital Transformation, OECD, 08 April 2022. https://www.oecd-ilibrary.org/development/development-co-operation-report-2021_63ebb18e-en. *See also:* Michael Pisa , Pam Dixon , Benno Ndulu and Ugonma Nwankwo, *Governing Data for Development: Trends, Challenges, and Opportunities*, Center for Global Development, November 12, 2020. https://www.cgdev.org/publication/governing-data-development-trends-challenges-and-opportunities .

[46] *Data Privacy, Ethics, and Protection: Guidance note on big data for achievement of the 2030 agenda,* United Nations Development Group, 2017. https://unsdg.un.org/sites/default/files/UNDG_BigData_final_web.pdf.

[47] *A World That Counts, Mobilizing the data revolution for sustainable development.* Prepared at the request of the United Nations Secretary-General by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. November 2014. p. 6. https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

Regarding data protection regulations in the development context, developing countries have been adopting national data protection legislation at a rapid pace. Much of this regulation is very similar to the GDPR in language and overall structure. While the advancement of data governance and protection is positive, it also means that standards efforts in all jurisdictions need to find the most effective ways of navigating the intersections of general data policy, national data protection legislation, the SDGs, AI impacts and context, and other contextual issues. There is much potential for benefit if more work regarding this specific intersection of SDGs, Data Protection Authorities, National Statistical Organizations, and local civil society and other groups can shed light on best practices that are fit for purpose and the country-level context. It is a challenging nexus, and there is much that could still be learned. Please also see the discussion regarding the need to address indigenous contexts in all AI work in the development and global context.

Recommendations:

1. Ensure that AI standards development work in the development context include the SDGs as a meaningful component of the work.
2. Ensure that there are guardrails in place to specifically prevent a “standards makers - standards takers” dichotomy.
3. Ensure that robust analysis of data protection legislation is conducted prior to the commencement of work, with ongoing checks. For example, a standard development process that facilitates web scraping for AI analysis in the development context should not be facilitated in jurisdictions where the data protection law prohibits such activity.
4. Ensure that all voices are heard, and that there is financial and other support to facilitate this in development contexts.

C. Global Standardization Efforts in AI and the Need to Address Indigenous Contextual Differences regarding Collective Privacy Rights and Data Sovereignty Approaches

The development of AI approaches can be at odds with indigenous socio-technical approaches. Ideologies and concepts of data, privacy, and ethics relating to modern indigenous peoples as articulated in national legal frameworks, tribal and governmental frameworks and agreements, and International customary law are important to take into account when developing global AI development programs, including standards development. This is precisely the moment in time to ensure that this happens.

While the conception of privacy and other rights is today primarily articulated as an individual rights in terms of legislation today, conceptions of privacy as a collective or community-based privacy right exist as well, and can be found articulated throughout the governance spectrum, from multilateral to national to tribal. International Customary Law provides significant indigenous rights to privacy and data sovereignty, primarily via the United Nations Declaration on the Rights of Indigenous Peoples, (UNDRIP), which

sets forth core rights of indigenous peoples' to govern themselves.[48] This would also apply to AI governance. Recently, the OECD has articulated the contours of AI harms in its Expert Group on AI Incidents.[49] The working definition of AI Incidents now includes impacts on groups of people and communities, not just individuals.

In national legislation, these ideas are set out in for example, the U.S. Federal Indian Law, Canadian law, and New Zealand law, among others. And finally, there is a critically important policy literature written by indigenous peoples regarding data held at the tribal level. Tsosie argues persuasively that tribal governments possess the authority to enact data privacy laws at the tribal level, and that this would help define what constitutes "tribal data." [50] The boundaries of what is and is not tribal data is a central question in the development of AI standards. Also, another issue exists, which is the idea of collective data ownership, and collective privacy rights, as well as the collective application of ethical principles. These types of approaches can be seen, for example, in the U.S. Indigenous Data Sovereignty Network and the Māori Data Governance Model, Te Kāhui Raraunga, among other indigenous governance frameworks, such as the First Nations Principles of OCAP. [51] OCAP, (Ownership, Control, Access, and Possession) for example, expressly establishes how First Nations' data and information in Canada will be collected, protected, used, or shared. Any AI standards development work in Canada should ensure that, for example, the OCAP principles are respected, and that representatives from Canada's First Nations are present for standards development processes.

In regards to AI specifically, the Māori have crafted a critically important policy literature, in which Kukutai *et al* explain that indigenous concepts of privacy are inherently collective. The New Zealand government is specifically working with the Maori to co-develop AI policy frameworks that are to be used whenever indigenous data or rights may be involved. New Zealand's approach to AI sets an important precedent, and WPF urges NIST to consider the structure of the New Zealand approach in its global standardization efforts. Here, a very brief background.

---

[48] United Nations Declaration on the Rights of Indigenous Peoples, Resolution adopted by the General Assembly on 13 September 2007, 62/295. UNDRIP https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_E_web.pdf

[49] *Expert Group on AI Incidents*, OECD. https://oecd.ai/en/network-of-experts/working-group/10836. See also: Stocktaking for the development of an AI incident definition, OECD, 27 October 2023. https://www.oecd.org/fr/publications/stocktaking-for-the-development-of-an-ai-incident-definition-c323ac71-en.htm. The most recent definitions of AI Incidents and impacts are published

[50] Tsosie, *Tribal Data Governance and Informational Privacy: Constructing 'Indigenous Data Sovereignty*,' 80 Montana Law Review 229 (2019)

[51] *The First Nations Principles of OCAP,* First Nations Information Governance Centre. https://fnigc.ca/ocap-training/

First, by way of background, New Zealand started early in its work on AI. In 2017, it established a Government Chief Data Steward (GCDS) role via mandate, and as such has already had time to produce a body of work and practice regarding data stewardship.[52] The Chief Data Steward is role is filled by the Chief Executive of Statistics New Zealand, or Stats New Zealand. The role has several functions: to set mandatory standards, to enable a "common approach to the collection, management and use of data across government," and to "direct the adoption of common data capabilities."

Notable work the Government Chief Data Steward has accomplished includes the development of a Data Strategy and Roadmap,[53] leadership in developing transparency and accountability for AI in the government context,[54] the development of a broad Data Stewardship Framework, and work on open data, and the development of a cooperative framework developed collaboratively with the Māori.[55] This work initially involved ensuring that work done regarding Covid-19 was respectful to the Maori approaches. Subsequently, this work was extended further in AI and accountability and standards development processes in collaboration with the Maori.

Structurally, New Zealand's framework of data stewardship is inclusive and interdependent across the whole of government. New Zealand describes its data stewardship framework as including a range of roles with governance functions in New Zealand's data system, including the:

- Government Chief Data Steward,
- Government Chief Information Security Officer,
- Government Chief Digital Officer,
- Government Chief Privacy Officer,

The Privacy Commissioner, Ombudsman, Auditor General, and Chief Archivist also have roles.

---

[52] *Government Chief Data Steward Mandate*, Office of the Minister of Statistics New Zealand. https://www.stats.govt.nz/assets/Uploads/Corporate/Cabinet-papers/Strengthening-data-leadership-across-government-to-enable-more-effective-public-services/strengthening-data-leadership-across-government-to-enable-more-effective-public-services-redacted.pdf.

[53] *The Government Data Strategy and Roadmap,* Government Chief Data Steward, September 2021. https://www.data.govt.nz/leadership/strategy-and-roadmap/.

[54] *Algorithm Assessment Report*, Stats NZ, 2018. https://www.data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-assessment-report/.

[55] *Māori Data Governance Co-design Review*, Te Kāhui Raraunga, January 2021. https://www.kahuiraraunga.io/_files/ugd/b8e45c_0b1a378da21c459eb4fb88dfbf6aea81.pdf . *See also*: *COVID-19 Lessons Learnt: recommendations for improving the resilience of New Zealand's government data system.* Stats NZ Tatauranga Aotearoa, March 2021. https://data.govt.nz/docs/covid-19-recs-report/.

The Privacy Commissioner's role is defined in the NZ Privacy Act of 2020, which has 13 information privacy principles, and requires agencies to report certain data breaches to the Privacy Commissioner. New Zealand's privacy laws are aware of GDPR, and as such it qualifies as a modern data protection law, but the Act is not identical to GDPR and uses different terminologies.

New Zealand's approach to algorithms, or AI and machine learning, as mentioned, has been progressive and inclusive. In 2018, New Zealand released its *Algorithm Assessment report*, which covered the practices of 14 government agencies.[56] It is among the earliest instances of a robust, mature discussion of data governance, management, standards, stewardship, open data, and privacy in the area of government use of algorithms. The 2018 report led to the July 2020 release of the first iteration of the *Algorithm Charter for Aotearoa New Zealand* by the Minister of Statistics.[57] The Charter is notable for its approach to providing for means of appeal of decisions informed by AI. New Zealand also released an initial algorithm toolkit in 2021 to implement the charter.[58]

As of 2024, the government of New Zealand has updated and expanded its AI-related materials in regards to its charter in an overarching toolkit, with its most recent update being 2023.[59] There are many features of the toolkit that are worth imitating, including the impressive list of signatories to the charter. These signatories specifically include the Ministry of Māori Development as well as other NZ Ministries.

Specific to indigenous-informed approaches to AI is the New Zealand Government's Algorithm impact assessment user guide, published in December 2023.[60] Beginning on page 29 of the Guide is a detailed discussion of New Zealand's relationship with the

---

[56] *Algorithm Assessment Report*, Stats NZ, 2018. https://www.data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-assessment-report/.

[57] *Algorithm Charter for Aotearoa New Zealand*, Stats NZ. July 2020. https://www.data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf.

[58] Government Algorithm Transparency and Accountability, Stats NZ. March 2021. https://www.data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability.

[59] Algorithm Charter for Aotearoa New Zealand, which includes foundational work from the following:
Principles for the safe and effective use of data and analytics
Government use of artificial intelligence in New Zealand [PDF 1.3 MB]
Trustworthy AI in Aotearoa - AI principles
Open government partnership
Data protection and use policy
and Privacy, human rights, and ethics framework [PDF 258 KB]

[60] *Algorithm impact assessment user guide*, New Zealand Government, Te Kāwanatanga o Aotearoa, December 2023. https://data.govt.nz/assets/data-ethics/algorithm/AIA-user-guide.pdf .

Māori and reflects with specificity its commitment to honor the Māori approach to data, and ensure the use of algorithms is consistent with the articles and provisions in its charter.

The guide notes on p. 29:

> **“General guidance**
>
> To meet the Partnership commitment in the Charter you should:
>
> - incorporate te ao Māori perspectives into the design and use of algorithms
> - ensure algorithm development and use is consistent with Te Tiriti o Waitangi
> - consider how Mäori data sovereignty will be maintained
> - assess how algorithm use will impact iwi and Mäori.
>
> Te ao Mäori acknowledges the interconnectedness and interrelationship of all living and non-living things via spiritual, cognitive, and physical lenses. This holistic approach seeks to understand the whole environment, not just parts of it. (This definition comes from Treaty of Waitangi/Te Tiriti and Māori Ethics Guidelines for: Al, Algorithms, Data and IOT.)”

Further into the Algorithmic assessment user guide, Question 5.3 on page 32 notes that:

> “**Māori data** is not owned by any one individual, but is owned collectively by one or more whanau, hapu or iwi. Individuals' rights (including privacy rights), risks and benefits in relation to data need to be balanced with those of the groups of which they are a part. (This definition comes from https:// www.temanararaunga.maori.nz/)
>
> **Mäori data sovereignty** recognises that Mäori data should be subject to Mãori governance — the right of Mäori to own, control, access and possess Mãori data. Mãori data sovereignty supports tribal sovereignty and the realisation of Mäori and iwi aspirations. (This definition comes from https: // www.temanararaunga.maori.nz/)”

We note that the express acknowledgement of indigenous approaches to data and AI by the government of New Zealand in its AI policy sets a critically important example. It is possible to incorporate multiple points of view regarding data. It will be important to ensure that global standards development efforts take note of the indigenous approaches that are either formal guidance or law in other countries.

Recommendations:

In regards to AI standards and policy in the indigenous context, it will be important for NIST to take into account indigenous approaches and national and other laws and agreements supporting those approaches. For example,

1. Any global standardization effort in AI must take into account and specifically address indigenous contexts and policies, understanding that these policies may differ substantially depending on regional, national, or subnational context. This is especially important for countries that are signatories to UNDRIP.
2. Data collection, analysis, and use relative to AI efforts must be conducted cooperatively and in a non-extractive manner in global contexts, and must also ensure that where indigenous contexts and approaches exist, that these are respected in regards to data and use of data.
3. There must be robust indigenous representation in the standardization development processes. WPF notes that there is a large and varied global indigenous context, and this has not been taken into account in most modern approaches to privacy, ethics, or other activities in many if not most cases regarding technology and data development, which often include AI.
4. A number of country-level governments have adopted UNDRIP as a matter of national law. For example, New Zealand is a signatory to UNDRIP and has formal agreements, and Canada has for example, passed an Act respecting the United Nations Declaration on the Rights of Indigenous Peoples, Bill C-15, which passed Canada's senate on June 16, 2021, and received royal assent on June 21, 2021 to become law.[61] This bill brings Canadian law into alignment with UNDRIP.

IV. Conclusion

The World Privacy Forum appreciates the opportunity to provide comments on NISTs work toward implementing the AI Executive Order. Thank you for your work, and WPF stands ready to assist with creating a robust evaluative environment for AI. We look forward to working co-operatively with NIST and multiple partners, countries, and etc. to craft thoughtful, ethical policy that reflects an empirically based, inclusive, and sound basis for AI standards, policies, and frameworks.

Respectfully submitted,

Pam Dixon
Executive Director,
World Privacy Forum

Documents included in this submission:

1. Comments of the World Privacy Forum regarding National Institute of Standards and Technology (NIST) Request for Information (RFI) Related to NIST's Assignments

[61] An Act respecting the United Nations Declaration on the Rights of Indigenous Peoples, Bill C-15, Parliament of Canada, https://www.parl.ca/LegisInfo/en/bill/43-2/C-15 .

Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence (Sections 4.1, 4.5, and 11), Docket Number: 231218-0309 (PDF)
2. Report: *Risky Analysis: Assessing and Improving AI Governance Tools An international review of AI Governance Tools and suggestions for pathways forward,* World Privacy Forum, 15 December 2023. (PDF)
3. Spreadsheet: WPF AI Governance Tools, Data only (PDF)

Some of the work and underlying thought included in this document is based on prior research and collaborative WPF work that included the following organizations and people. The relevant work, when present, is cited in the document footnotes.

- UN Statistical Commission WG on Data Stewardship, 2021 - 2024
- OECD AIGO Expert Group on AI Incidents
- The Center for Global Development
- Kate Kaye, Michael Pisa, Benno Ndulu, Ugonma Nwankwo, Robert Gellman.
- Research: Pam Dixon, with additional research from Kate Kaye regarding content provenance.
- Section I of these comments is based on extensive co-authored original research by Kate Kaye and Pam Dixon from *Risky Analysis: Assessing and Improving AI Governance Tools An international review of AI Governance Tools and suggestions for pathways forward,* World Privacy Forum, 15 December 2023. A dataset of AI governance tools is included in this submission. Due to page limitations, the dataset has been included as a separate document.