Microsoft's Response to the Request for Information of the National Institute of Standards and Technology

Developing Artificial Intelligence Standards

I. Introduction and Executive Summary

Microsoft Corporation ("Microsoft") welcomes the opportunity to provide comments to the National Institute of Standards and Technology ("NIST") in response to its May 1, 2019, request for information ("RFI").[1] The RFI seeks comments on the current state, plans, challenges, and opportunities regarding the development and availability of artificial intelligence ("AI") technical standards and related tools, as well as priority areas for federal involvement in AI standards-related activities. Microsoft supports NIST's efforts to meet its obligations under the February 11, 2019, Executive Order on Maintaining American Leadership in Artificial Intelligence ("Executive Order on AI"), including driving development of appropriate technical standards that "enable the creation of new AI-related industries and the adoption of AI by today's industries."[2]

AI has the potential to transform economies and to address societal challenges by enhancing human decision-making processes with additional insights and intelligence, leading to better outcomes and improvements in every aspect of people's lives. Like many emerging technologies, however, AI can be misused in ways that can create individual, organizational, and societal concerns. As a leading supplier of AI solutions, Microsoft is committed to responsible AI innovation. We recognize the need for sustained, constructive engagement among industry, government, academia, and other stakeholders in developing and using AI. We recommend policymakers in the United States to commit to high-level ethical and moral principles necessary to support the development and use of AI in a manner that maximizes the benefits of the technology while protecting individuals and society from its potential risks. In addition, we encourage the U.S. Government broadly and NIST specifically to support ongoing efforts to develop foundational standards for AI technologies, which can help businesses adopt practices consistent with such high-level principles.

Microsoft recognizes the benefits of AI cannot be realized unless AI is deemed trustworthy by individuals and society. We believe the starting point for creating trust in AI is taking a human-centered approach, and grounding system designs in universal, timeless values shared by stakeholders from industry, government, civil society and the research community. In September 2016, we took a first step toward these goals by co-founding the Partnership on AI ("PAI"), 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."[3] Microsoft was also a strong supporter of the International Standards Organization's establishment in 2017 of a subcommittee

---

[1] Department of Commerce, National Institute for Standards and Technology, Notice; Request for Information (May 1, 2019), https://www.govinfo.gov/content/pkg/FR-2019-05-01/pdf/2019-08818.pdf.

[2] Exec. Order No. 13859, 84 Fed. Reg. 3967, 3967 (2019), https://www.govinfo.gov/content/pkg/FR-2019-02-14/pdf/2019-02544.pdf (the "Executive Order on AI").

[3] Partnership on AI, https://www.partnershiponai.org.

(continued…)

on artificial intelligence within its technical committee on information technology.[4] In January 2018, we shared our views on AI in the book The Future Computed: Artificial Intelligence and Its Role in Society,[5] and we have identified six principles that we believe should guide the cross-disciplinary development and use of AI: fairness, reliability and safety, privacy and security, inclusivity, transparency, and accountability.[6] We also strive to ensure our company reflects these principles, including by creating a new internal body, the AI and Ethics in Engineering & Research ("AETHER") Committee, which advises our Senior Leadership Team on policies, processes and best practices on issues of AI and ethics.

Microsoft recommends that the NIST and the U.S. Government adopt three priorities in engaging on AI standardization:

- *First*, policymakers in the U.S. Government should reinforce their commitment to core principles that should guide the United States in supporting the development and use of responsible AI. The United States has already adopted a series of principles developed by the Organisation for Economic Co-operation and Development ("OECD"), as described in Section IV, and NIST should recognize the importance of such principles in the work it does supporting the development of both standards and tools for AI.

- *Second*, NIST and U.S. Government agencies should engage in existing standards bodies or projects when it invests in AI standardization. NIST has the opportunity to independently assemble stakeholders from government, academia, and industry to produce a gap analysis of needed standardization or to identify areas that may benefit from future standardization work. Yet NIST should avoid becoming a standards-setting organization in seeking to address those gaps.

- *Third*, NIST should increase its involvement in efforts to build out new tools and best practices relating to AI. Specifically, NIST should support the development of such tools and services in connection with AI used to deliver government services.

In working toward these goals, the U.S. Government should recognize the role of standards in achieving policy outcomes. As the Executive Order on AI states, the development of technical standards plays an important role in "shaping the global evolution of AI in a manner consistent with our Nation's values, policies, and priorities."[7] The Government's efforts should recognize the foundational role played by standards generally, and by international standards specifically, in supporting high-level principles that foster trustworthy AI.

---

[4] *See* International Organization for Standardization*, ISO/IEC JTC 1/SC 42: Artificial Intelligence,* https://www.iso.org/committee/6794475.html; JTC1, *ISO/IEC JTC 1/SC 42 Artificial Intelligence* (last revised Mar. 2019), https://jtc1info.org/sd_2-history_of_jtc1/jtc1-subcommittees/sc-42/.

[5] Brad Smith & Hary Shum, *The Future Computed: Artificial Intelligence and Its Role in Society* (Jan. 17, 2018), https://blogs.microsoft.com/blog/2018/01/17/future-computed-artificial-intelligence-role-society.

[6] Microsoft, *Six Principles for Developing and Deploying Facial Recognition Technology* (Dec. 2018), https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2018/12/MSFT-Principles-on-Facial-Recognition.pdf.

[7] Executive Order on AI at 3967.

II. The Role of Standards in Supporting Trustworthy AI

A. Standards in a Broader Context

Ethical and moral principles are the foundations upon which nations establish laws, regulations, and policy. Standards, in turn, are a tool that help industry to implement policy or legal requirements or help to demonstrate adherence to those requirements. As a result, any work to develop standards for a new technology must take place in the context of laws and regulations affecting that technology.

Standards cannot replace the crucial role of policy and law in identifying the ethical and moral principles relevant to AI. Microsoft believes that respect for societal norms and sovereign authority are essential to developing trustworthy AI. The data scientists and engineers who build AI technologies should not define these norms for the world. Rather, society and government must do so, including through ethical and moral principles, laws, regulations, and policy. Data scientists and engineers developing AI must then take those norms into account when developing and deploying new technologies.

AI technology is already governed by many national, regional, and sector-specific laws and regulations. For example, laws that protect consumers from predatory home loans still apply to lenders that incorporate AI-enabled models into their lending practices. Similarly, health and medical regulations still apply to doctors that deliver healthcare using AI technologies. In most cases, existing laws are capable of addressing issues raised by the use of AI. It is only when the use of AI raise concerns not addressed by an existing legal framework that new guidance should be considered. In these scenarios, policymakers must thoughtfully address concerns raised by AI in line with the core high-level principles supporting the responsible use of AI. In doing so, policymakers should consult with industry, academic, government, and other stakeholders, to ensure any response to those concerns does not inhibit the responsible use and deployment of AI.

Standards are only one tool among many that can help to achieve compliance with laws, regulations, and principles that reflect societal norms. Other tools may include open source software ("OSS"), codes of conduct, self-attestation, and operational guidelines. Rather than focus on any one tool, such as standards, we encourage NIST to maintain a holistic view of these mechanisms to determine which is appropriate for implementing any desired outcome. However, the introduction of new regulations and policies has created an increased need to ensure compliance with certain norms. In this context, standards are often critical in supporting assurance practices that assess compliance for norms required by law or policy. NIST itself has done excellent work in producing guidance around conformity assessment which further clarifies this point.[8]

B. The Standards Landscape

Organizations are already developing AI standards, including cooperative, industry-led efforts that benefit from subject matter experts representing industry, government, academia and civil society, as discussed in Part III. For purposes of this discussion, standards can be grouped into three categories:

[8] Department of Commerce, National Institute for Standards and Technology, NIST Special Publication 2000-01: *ABC's of Conformity Assessment* (2018), *available at* https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.2000-01.pdf.

- *Foundational Standards.* These standards establish globally shared basic concepts, such as terminology, use cases, and reference architectures. This type of standardization work creates a common understanding upon which trustworthy AI policy and practices can be built.

- *Technical Interoperability Standards*. These standards typically establish mechanisms such as formats, protocols or interfaces that enable disparate systems to communicate.

- *Management Standards*. These standards establish governance guidelines and controls-based operational processes that help to form the criteria for responsible behavior and enable organizations to demonstrate conformity to best practices and regulations.

In addition to standards, other assurance mechanisms can be used to determine if a technical specification or practice is achieving compliance with a particular norm. Such mechanisms include voluntary implementation of standards, contracts, attestation, and risk management.

Development of Standards. Standards are most useful when they are developed through a transparent process, with open participation, and reflect a bottom-up approach to development. Healthy standardization of foundational concepts and management practices also discourages the use of standards as a discriminatory barrier to market access. While different countries will apply their own ethical and legal regimes to AI, all countries benefit from promoting economic growth and augmenting human capabilities with AI technologies. That growth can be disrupted if conflicting or discriminatory standards are implemented on a country-by-country or regional basis. Moreover, to the extent that standards are used to assess conformity with laws and regulations, they should be developed in organizations that adhere to the principles in the World Trade Organization's Technical Barriers to Trade Agreement ("WTO-TBT").[9] For the U.S. Government, this is consistent with the principles of OMB Circular A.119, which recognizes that agencies should consider international standards, rather than domestic standards, consistent with interests in promoting trade and implementing the provisions of international treaty agreements.[10] This approach also further encourages participation from industry, academia and government.

Changing Role of Standards. Standardization strategies should also be developed in a manner that reflects the changing landscape around how standards are used today. Historically, interoperability standards risk creating innovation dead-zones rather than promoting market opportunities if standardization is undertaken too early in the maturity cycle of a new technology domain. Yet there is no doubt that technical interoperability standards can act as market-makers, and are generally considered desirable, pro-competitive industry activities. However, as use of new collaborative development methodologies have emerged over the past two decades, the traditional approach to standardization has been disrupted. The global trend of creating open source software is now frequently used across sectors as another means to achieve the same objectives of market-makers and interoperability between disparate systems. Both OSS and traditional standards have thus become

[9] World Trade Organization, *Agreement on Technical Barriers to Trade,* https://www.wto.org/english/docs_e/legal_e/17-tbt_e.htm.

[10] Office of Management and Budget, Circular No. A-119 Revised (Feb. 10, 1998), *available at* https://www.whitehouse.gov/wp-content/uploads/2017/11/Circular-119-1.pdf ("OMB Circular No. A-119").

legitimate pathways for promoting market opportunities for all market players and any government policy or standardization strategy should take this trend into consideration.

At the same time, there is a broader societal need for trust and accountability for the organizations that produce or use software technologies—and it is in this context where AI standardization has an essential role to play, as opposed to OSS technology solutions. The last decade has seen the introduction of regulations and policies that address cybersecurity, privacy and the movement and use of data, all of which are critical to AI. Standards that enable companies to meet these regulatory and policy requirements through the establishment of globally accepted conformance criteria can help drive trust and accountability in AI technologies.

III. Development of AI Technical Standards and Related Tools: Status and Plans

In the past two years, organizations have begun a range of projects aimed at producing standards and tools relevant to AI. We set out below a partial list of those efforts, which complement the development by policymakers of high-level principles and values to support trustworthy AI.

A. AI Standards

International and national organizations are already working to develop both standards specific to AI and broader standards that address core concerns raised by AI.

International Standards. A small number of international standards bodies are working on AI, including:

- ISO / IEC JTC 1 SC 42 – Artificial Intelligence.[11] Twenty-six countries are already participating in SC 42's work, and an additional 12 countries have an observer status. The committee's scope is broad and encompasses standardization in the area of AI. Currently, it is focused on: (1) foundational standards[12] that will define the bases of AI internationally, and (2) creating trustworthiness through risk management[13] and AI governance.[14] SC 42 has already released three Big Data standards and its first AI publication, titled an "Overview of Trustworthiness in Artificial Intelligence"[15] is expected by the end of this year. Microsoft welcomes and supports the current efforts of SC 42, which benefits from an active U.S. delegation that is well-positioned to shape its efforts in developing an international framework.

[11] International Organization for Standardization*, ISO/IEC JTC 1/SC 42: Artificial Intelligence,* https://www.iso.org/committee/6794475.html.

[12] JTC1, *ISO/IEC JTC 1/SC 42 Artificial Intelligence* (last revised Mar. 2019), https://jtc1info.org/sd_2-history_of_jtc1/jtc1-subcommittees/sc-42/ (listing as a working group "Foundational Standards").

[13] *Id.* (listing as a trustworthiness project "ISO/IEC NP 23894: Information technology – Artificial Intelligence – Risk Management").

[14] *Id.* (listing as a joint working group "Governance Implications of AI").

[15] International Organization for Standardization, *ISO/IEC PDTR 24028: AI – Overview of Trustworthiness in Artificial Intelligence* (forthcoming 2019), https://www.iso.org/standard/77608.html?browse=tc.

(continued…)

- ISO/IEC JTC 1 SC 38 — Cloud Computing and Distributed Platforms.[16] This committee is focused on cloud computing and distributed platforms, which are not specific to AI but are critical to effectively using AI. The committee's efforts on data taxonomy and data sharing, which are already available, are a key resource for companies engaging in machine learning.[17] The committee is also preparing a data-sharing standard,[18] which will be important in developing the AI ecosystem and supporting the digital economy through sharing data and information. That new standard is expected to be published in 2022 and Microsoft supports these efforts.

- ISO /IEC JTC 1 SC 27 — Information Security, Cybersecurity, and Privacy Protection.[19] This committee is a world reference on cybersecurity and privacy, two issues that are broadly relevant but are of particular importance to AI. The committee is focused specifically on issues including information security management systems, privacy information management systems, cryptography, and Big Data security and privacy.[20] It also has conducted initial investigations into privacy for AI, and those efforts remain in progress. Microsoft supports these efforts and welcomes the committee's study of how AI will affect these domains.

- IEEE C/S2ESC – Software & Systems Engineering Standards Committee P7000-series.[21] This organization is also conducting work relevant to AI standards. In 2016, IEEE started the Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, which led to the launch of standards projects that remain under development. Microsoft has not been actively engaged in these projects but notes that the committee has announced standards covering a large number of subject areas, including ethics, transparency, bias, and autonomous systems.

National and Regional Standards. Outside of these international efforts, countries and regions are also engaging in their own activities to develop AI standards, which will establish regulatory and policy frameworks to address trustworthy AI within the context of their own legal and ethical systems. For example, last month the European standards organization CEN CENELEC launched a focus group to

---

[16] International Organization for Standardization*, ISO/IEC JTC 1/SC 27: Information Security, Cybersecurity and Privacy Protection*, https://www.iso.org/committee/601355.html.

[17] *See, e.g.*, International Organization for Standardization*, ISO/IEC DIS 22624: Taxonomy Based Data Handling for Cloud Services* (2019), https://www.iso.org/standard/73614.html; International Organization for Standardization, *ISO/IEC 19944: Data Flow, Data Categories and Data Use* (2017), https://www.iso.org/standard/66674.html.

[18] International Organization for Standardization, *ISO/IEC AWI 23751: Data Sharing Agreement (DSA) Framework* (forthcoming 2022), https://www.iso.org/standard/76834.html?browse=tc.

[19] International Organization for Standardization*, ISO/IEC JTC 1/SC 27: Information Security, Cybersecurity and Privacy Protection*, https://www.iso.org/committee/45306.html.

[20] *See, e.g.*, International Organization for Standardization, *ISO/IEC 20889: Privacy Enhancing Data-Identification Terminology and Classification of Techniques* (2018), https://www.iso.org/standard/69373.html.

[21] IEEE, https://standards.ieee.org/project/7000.html.

(continued…)

conduct a one-year study on the need to start projects on AI.[22] In China, the government encourages Chinese standard authorities and AI companies to develop AI standards for promoting China AI industry. We are aware of eight China AI standard organizations (SDOs), that develop different kinds of standards, including AI national standards, social standards, and industry standards that are extensive, including algorithms, AI platforms, speech recognition, and other technologies.[23] In Canada, a domestic organization, the Chief Information Officer Strategy Council, provides a forum for the country's chief information officers to collectively focus on influencing the Canadian information and technology ecosystem. As part of its standards pillar, the organization is developing two standards, including one focused on the ethical design and use of automated decision systems.[24]

Sector-Specific Standards. Given the continued development and innovation of AI technology and the ongoing work to develop a foundational set of horizontal standards for the technology overall, it is premature to devote significant resources to developing sector-specific vertical standards at this time. Sector-specific standards should not be developed unless they are driven by market or regulatory needs and are informed by sufficient implementation experience. Once the baseline international standards are established for AI, those standards may be tailored to create sector-specific standards suited to a range of industries and geopolitical regimes. In turn, such sector-specific standards may eventually benefit broader AI standards—such as by allowing for more precise identification of the appropriate criteria to evaluate an AI system, which may vary based on the industry in which it is used. Examples of the benefits of sector-specific efforts are the ITU-T Focus Group on artificial intelligence for health ("AI4H"), which was established in July 2018 and is working with the World Health Organization to establish a standardized assessment framework for evaluating AI-based methods for health, diagnosis, triage and treatment decisions,[25] and the ITU-T Focus Group on Machine Learning for Future Networks including 5G, which was re-established in March 2019 and is analyzing the impact of the adaption of ML for future networks.[26]

B. AI Tools

In addition to standards, a broad range of tools can help to establish a common understanding of the issues raised by AI, provide practical solutions, and evaluate approaches to facilitate trustworthy AI. Microsoft has continually invested in the research and development of such tools for new technologies, including AI. These tools allow for gaining and sharing practical experience, which is a prerequisite to

[22] *See* CEN-CENELEC, *Kick-Off Meeting of the CEN-CENELEC Focus Group on Artificial Intelligence on 24 April*, https://www.cencenelec.eu/news/brief_news/Pages/TN-2019-018.aspx.

[23] These organizations include the SAC AI Standard General WG, TC28 SC42(to be established), CESA(China Electronics Standardization Association), AIIA(AI Industry Alliance), AIIIA(China AI Industry Innovation Alliance), CCSA TC1 WG1, AIOSS(China AI Open Source Software Development Alliance) and AITISA(AI Industry Technology Innovation Strategic Alliance).

[24] *See* CIO Strategy Council, Notice of Intent (July 31, 2018), https://ciostrategycouncil.com/standards/new-projects/.

[25] *See* ITU Telecommunications Standardization Sector, Focus Group on Artificial Intelligence for Health, *FG-AI4H*, https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx.

[26] *See* ITU Telecommunications Standardization Sector, Focus Grou pon Machine Learning for Future Networks Including 5G, FG-ML5G, https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx.

(continued…)

developing more advanced technical standards, such as emerging evaluation approaches, metrics, and interoperability.

Microsoft believes that AI systems with consequential impact and/or those that pose high risk should be transparent and explainable. Microsoft has demonstrated this through a facial recognition transparency note[27] and supports the use of risk management practices to enable companies to detect potential issues and remedy them. Moreover, tools that foster more responsible development and use of AI should be open and interoperable.

Below is a partial list of Microsoft's publicly-available tools and publications, which fall into three areas: (1) business understanding tools, which include guidelines and best practices to support customers' and partners' decision-making processes in AI systems; (2) data acquisition and understanding tools, which help developers, integrators, and users of AI systems to understand and address common challenges related to training data for machine learning models, and (3) modeling tools, which share Microsoft's findings in the area of AI systems and machine learning algorithms intelligibility.

Business Understanding Tools

- *Human-AI design principles* (research paper).[28]
- *Conversational AI guidelines* (research paper and AI school course).[29]

Data Acquisition & Understanding

- *Datasheets for datasets* (research paper)[30]
- *Homomorphic encryption* (OSS)[31]
- *Differential privacy* (used for telemetry)[32]

[27] Microsoft AI*, Transparency Note: Azure Cognitive Services: Face API* (last updated Mar. 29, 2019), https://azure.microsoft.com/mediahandler/files/resourcefiles/transparency-note-azure-cognitive-services-face-api/Face%20API%20Transparency%20Note%20(March%202019).pdf.

[28] Saleema Amershi, Mihaela Vorvoreanu, and Eric Horvitz, *Guidelines for Human-AI Interaction Design* (Feb. 1, 2019), https://www.microsoft.com/en-us/research/blog/guidelines-for-human-ai-interaction-design/.

[29] Microsoft Corporation, *Responsible Bots: 10 Guidelines for Developers of Conversational AI* (Nov. 2018), https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf (research paper); Responsible Conversational AI, https://aischool.microsoft.com/en-us/responsible-ai/learning-paths/responsible-conversational-ai (last visited May 29, 2019) (school course).

[30] Timnit Gebru, *et al.*, *Datasheets for Datasets* (last revised Apr. 14, 2019), https://arxiv.org/abs/1803.09010.

[31] Microsoft SEAL, https://www.microsoft.com/en-us/research/project/microsoft-seal/ (last visited May 29, 2019).

[32] Project Laplace, https://www.microsoft.com/en-us/research/project/project-laplace.

(continued…)

- *Secure MPC* (part of Azure Confidential Computing)[33]

Modeling

- *InterpretML Library* (OSS in Build 2019)[34]
- *Reductions approach to fair classification* (research paper and related Github repo)[35]

C. Alternate Methods

As described in Section II.B, there is a significant trend toward the use of open source software to provide software-based interoperability. In many cases, this OSS work has replaced traditional standardization activities and represents a market-relevant path to interoperability.[36] To the extent NIST considers the use of OSS as a viable path for interoperability, we respectfully suggest that it is careful not to recommend a single OSS project but rather the performance or outcome requirements that are appropriate regardless of which OSS project is used to meet the need. NIST should avoid technology mandates by recognizing that all OSS projects are technology-specific and thus care is needed to appreciate the dynamics that make it similar to but different from standardization.

IV. Defining and Achieving U.S. AI Technical Standards Leadership

At the broadest level, the U.S. can help to support standardization efforts by demonstrating a commitment to high-level principles needed to develop and use trustworthy AI. These ethical and moral principles should reflect the values needed to support trustworthy AI, consistent with the Executive Order's recognition that "[c]ontinued American leadership in AI is of paramount importance to . . . shaping the global evolution of AI in a manner consistent with our Nation's values, policies, and priorities."[37]

A. High-Level Principles on AI

Governments across the world have committed to articulating a series of ethical and moral principles that are foundational to trustworthy AI. Although these principles vary, there is a clear convergence around a core set of ethical and moral concerns. In January 2019, Singapore released a framework on the ethical and responsible use of AI, focused on two "high-level guiding principles that promote trust in

[33] Stefano Tempesta, *Secure Multi-Party Machine Learning with Azure Confidential Computing* (April 2019), https://www.microsoft.com/en-us/research/project/microsoft-seal/.

[34] Microsoft, Interpret ML Library, https://github.com/microsoft/interpret.

[35] Microsoft, FairLearn, https://github.com/microsoft/fairlearn.

[36] For example, ONNX is an open neural network exchange format being developed as a community project. See ONNX, https://onnx.ai/. Similarly, Khronos.org hosts the Neural Network Exchange Format ("NNEF"), a similar concept standard that is a specification-first, rather than code-first, approach. See Khronos Group, NNEF Overview, https://www.khronos.org/nnef.

[37] Executive Order on AI at 3967.

(continued…)

AI and understanding of the use of AI technologies."[38] Those principles support decision-making processes that[39] are explainable, transparent, and fair, and AI solutions that are "human-centric." In April 2019, the Australian Government sought comment on eight core principles for AI, focused on generating net benefits, doing no harm, regulatory and legal compliance, privacy protection, fairness, transparency and explainability, contestability, and accountability.[40] The same month, the High-Level Expert Group on AI set up by the European Commission published revised ethics guidelines that identified three components of trustworthy AI, stating the technology should be lawful, ethical, and robust.[41]

Microsoft recommends that the U.S. Government demonstrate its commitment to such core principles, which would build on its adoption last month of the OECD AI principles.[42] Under those principles:

1. AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.

2. AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.

3. There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.

4. AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.

5. Organizations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

---

[38] Singapore Personal Data Protection Commission, *A Proposed Model Artificial Intelligence Governance Framework* (January 2019), https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf

[39] *Id.*

[40] See Australian Government, Department of Industry, Innovation and Science, *Artificial Intelligence, Australia's Ethics Frameworks, A Discussion Paper* (DATE), https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf.

[41] High-Level Expert Group on Artificial Intelligence Set up by the European Commission, E*thics Guidelines for Trustworthy AI* (April 8, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

[42] *See* OECD, *Forty-Two Countries Adopt New OECD Principles on Artificial Intelligence* (May 22, 2019), https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm.

(continued…)

Just as governments are adopting AI ethical frameworks and principles, companies are also identifying ethical and value-based principles for engaging on AI. For example, Microsoft has adopted six ethical principles to guide the development and use of artificial intelligence.[43] They are:

- *Fairness*. AI systems should treat all people fairly.
- *Inclusiveness*. AI systems should empower everyone and engage people.
- *Reliability & Safety*. AI systems should perform reliably and safely.
- *Transparency*. AI systems should be understandable.
- *Privacy & Security*. AI systems should be secure and respect privacy.
- *Accountability*. AI systems should have algorithmic accountability.

At Microsoft, we apply these principles throughout our business, and have committed to using AI to address pressing societal issues, including our AI for Good initiative that uses AI to help the world recover from disasters, address the needs of children, protect refugees and displaced people, and promote human rights.[44] Our AI for Accessibility project also aims to accelerate the development of accessible and intelligent AI solutions to benefit 1 billion-plus people with disabilities around the world.[45] And our AI for Earth initiative has committed $50 million over 5 years to providing AI tools for researchers working on environmental challenges.[46]

Demonstrating a commitment to a core set of high-level ethical values and principles for AI is critical, because it enables organizations to act on those values and principles, including through development of policies and standards. Microsoft therefore recommends the United States demonstrate its commitment to a set of high-level principles to foster trustworthy AI.

B. Support Ongoing Standardization Efforts Relevant to High-Level Principles

The U.S. can complement this commitment to high-level principles for AI values and principles by supporting ongoing standardization work, without duplicating those existing efforts.

In particular, Microsoft encourages NIST to continue supporting existing efforts led by ISO/IEC JTC SC 42. Doing so would show that contributing to the development of robust international standards benefits everyone. Open, global standardization systems enable market competition to play out, and historically

---

[43] Microsoft, *Microsoft AI Principles*, https://www.microsoft.com/en-us/ai/our-approach-to-ai.

[44] Brad Smith, *Using AI to Help Save Lives* (Sept. 24, 2018), https://blogs.microsoft.com/on-the-issues/2018/09/24/using-ai-to-help-save-lives/.

[45] Brad Smith*, Using AI to Empower People with Disabilities* (May 7, 2018), https://blogs.microsoft.com/on-the-issues/2018/05/07/using-ai-to-empower-people-with-disabilities/.

[46] Paul Fleming, *On World Water Day, Microsoft is Delivering New Approaches to Ensure We Leave No One Behind* (Mar. 22, 2019), https://blogs.microsoft.com/on-the-issues/2019/03/22/on-world-water-day-microsoft-is-delivering-new-approaches-to-ensure-we-leave-no-one-behind/.

(continued…)

have led to growth in U.S. technology sectors. Countries that depart from such international efforts, and that instead adopt top-down approaches in which a government mandates the establishment of specific standards, have tended to fall behind in industry competitiveness.

As the United States has recognized, when international standards exist or their completion is imminent, standardizing bodies benefit from using those existing standards, except when they would be ineffective or inappropriate. This principle is embodied in the World Trade Organization's Agreement on Technical Barriers to Trade Agreement, which sets out in Annex 3 a code of good practice for the preparation, adoption, and application of standards.[47] As that code notes, good standardization practices focus on establishing responsible behavior for all actors, enabling borders to be open to the free flow of data, and establishing a level playing field for active competition—which is the underpinning of market economics. The United States has recognized the importance of this approach to standardization for more than 20 years. In 1998, the Office of Management and Budget ("OMB") issued Circular No. A-119, which implements the United States' commitments to support the WTO TBT Agreement.[48] Under that OMB guidance, when a "voluntary consensus standards body is in the process of developing or adopting a voluntary consensus standard that would likely be lawful and practical for an agency to use, and would likely be developed or adopted on a timely basis, an agency should not be developing its own government-unique standard and instead should be participating in the activities of the voluntary consensus standards body."[49] As a result, federal agencies are to use voluntary consensus standards in lieu of government-unique standards, unless doing so is inconsistent with law or impractical.

In addition to recognizing the need to *use* standards already under development, the United States has also recognized the need to *support* the development of international standards. Under OMB's guidance, agencies "must consult with voluntary consensus standards bodies, both domestic and international" and "must participate with such bodies in the development of voluntary consensus standards when consultation and participation is in the public interest and is compatible with their missions, authorities, priorities, and budget resources."[50] One example of this approach is NISTR 8074, which expressly notes that the U.S. should rely on international standards where possible in the context of cybersecurity, and avoid duplicative efforts.[51] Microsoft accordingly recommends that the U.S. government broadly and NIST specifically to continue to play a leadership role in supporting the ongoing development of existing international standards.

V. Prioritizing Federal Government Engagement in AI Standardization

U.S. efforts to support the development of AI standards should focus on identifying and filing gaps in existing standards, including those under development. In particular, Microsoft recommends that NIST

[47] WTO TBT at Annex 3.

[48] OMB Circular No. A-119.

[49] *Id*.

[50] *Id*.

[51] Department of Commerce, National Institute of Standards and Technology, NISTIR 8074 Vol. 2: *Supplemental Information for the Interagency Report on Strategic U.S. Government Engagement in International Standardization to Achieve U.S. Objectives for Cybersecurity* 2 (2015), https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8074v2.pdf.

(continued…)

lead a collaborative process to identify potential gaps in existing standards work, with the goal of addressing those gaps through efforts led by international standards organizations.

NIST is well positioned to undertake such an effort. It has been a positive participant in the early rounds of international standardization of AI and provides cross-departmental leadership for the U.S. in ensuring that agencies rely on existing and developing standards rather than create government-unique standards, in line with OMB's guidance in Circular A.119.[52] NIST has established credibility, thought leadership and has promoted the value of multi-stakeholder participation, including through the Cybersecurity Framework.

A. The Need to Prioritize and Address Gaps in Standardization Efforts

Standards organizations are already working on a number of efforts relating to AI. Microsoft recommends that NIST prioritize its engagement in the standardization of foundational standards for AI (including the AI lifecycle) within ISO/IEC JTC 1 and its SC 42 and complement those efforts by leading a process to identify gaps in existing standardization work with subject matter experts from government, industry and academia. As part of this effort, NIST could review standards relevant to areas of governance, cybersecurity, privacy, intelligibility, bias and fairness, data quality, and data sharing, to prioritize and identify potential gaps in these existing efforts.

These efforts should take place within the ongoing work of ISO/IEC JTC 1 SC 42 WG 1, which is scheduled to complete standardization of foundational standards this Fall. We recommend that NIST actively engage in the review and development of these standards to ensure their quality and usability in supporting trustworthy AI. Leading a gap analysis of AI standards efforts within ISO/IEC JTC 1 SC 42, rather than as a standalone NIST effort, would also leverage NIST's expertise to assist in the potential development of standards that could enable global solutions.

If NIST undertakes such a gap analysis, it should address at least the following areas:

- *Risk Management for AI.* Risk management is a preventative process that can help ensure that a specific AI product or service is trustworthy throughout its lifecycle. In Microsoft's view, mitigating risk is at the core of advancing trustworthy AI. Risk management practices are especially suited to new technologies where the unknown is greater than the known. Examples of such frameworks include ISO/IEC Technical Report 27103:2018 and the NIST Cybersecurity Framework.[53] Risk management for AI, in addition to security and privacy, should explicitly address concerns related to fairness, transparency, robustness and safety of AI systems and should focus not only on risks to an organization but also to third-party partners and suppliers, end users, and society more broadly. Creation of a controls-based risk management standard for AI would bring together important building blocks such as AI lifecycle, AI governance, as well as standards addressing cybersecurity, privacy, explainability and interpretability, bias, data quality, and data sharing. If a new controls-based, AI risk-management standard is to be

[52] OMB Circular No. A-119.

[53] International Organization for Standardization, *ISO/IEC TR 27103:2018: Cybersecurity and ISO and IEC Standards* (2018), https://www.iso.org/standard/72437.html; National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1 (Apr. 16, 2018), https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf.

created, it should be undertaken in an international standards organization, such as ISO/IEC JTC 1, and not as a standalone NIST activity.

- *AI Lifecycles*. In Microsoft's view, an AI lifecycle is the underlying backbone for applying an organization's risk management process to an AI system. A lifecycle framework can help organizations describe a particular AI system in accordance with the organization's role and the system's foreseeable uses. The risks and the mitigation measures identified through the risk management process would be mapped to and implemented throughout the system's lifecycle. Microsoft encourages NIST to engage in the definition of an AI lifecycle framework as a necessary tool for supporting trustworthy AI.

- *AI Governance.* Governance structures and data handling practices must be sensitive to context and should be tailored to individual scenarios in which AI is used. Governing bodies set out a purpose and other parameters for an organization—and are accountable for the whole organization. As a result, a governance standard for AI could identify parameters such as an organization's: (1) strategic decisions, including markets to serve and products and services to provide, (2) stakeholder engagement, to set investments, expected behaviors, culture and values, and (3) risk appetite, to outline the nature and extent of risks that the organization is willing to take on in the pursuit of its goals. Governance frameworks also put in place management infrastructure and processes to efficiently and effectively deliver the objectives of an organization within these parameters. These responsibilities could be implemented at the outset of a risk management process and involved later as appropriate.

- *Cybersecurity.* Risks and threats to and from an AI system may go beyond traditional cybersecurity threats and risks, to include risk sources such as bias, lack of transparency, and manipulation of external datasets. While the Cybersecurity Framework is widely used in industry and government, it may be appropriate to also identify AI-specific risks presented by a system and understand how an organization will manage those risks. One starting point for identifying core areas of cybersecurity standards that may be updated in light of AI is NISTR 8074, Interagency Report on Strategic U.S. Government Engagement in International Standardization to Achieve U.S. Objectives for Cybersecurity.[54] Just as NIST previously drafted a report on cybersecurity standards for the Internet of Things[55] it can undertake a similar report to assess AI cybersecurity standards. In future revisions of these references, Microsoft recommends NIST to undertake a process to involve stakeholders in considerations about future updates to address potential issues arising from AI.

- *Privacy*. AI is built on data, much of which is obtained from people. Processing that data through AI and machine learning creates a potential for large-scale privacy concerns, including because of the potential for inferred data to make associations about a person not possible

[54] Department of Commerce, National Institute of Standards and Technology, NISTIR 8074 Vol. 1: *Interagency Report on Strategic U.S. Government Engagement in International Standardization to Achieve U.S. Objectives for Cybersecurity* 2 (2015), https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8074v1.pdf.

[55] Department of Commerce, National Institute of Standards and Technology, NISTIR 8200: *Interagency Report on the Status of International Cybersecurity Standardization for the Internet of Things (IoT)* (2018), https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8200.pdf.

without AI. One potential roadmap for addressing data protection concerns is through mechanisms established in the EU's General Data Protection Regulation ("GDPR"), although any mechanism should be internationally-accepted and interoperable with other frameworks. In addition, incentives are needed to promote the use of additional privacy-protective measures, such as effective de-identification techniques, coupled with legal or administrative controls that penalize efforts to re-identify data, and balance the rights of data subjects and the public interest, which must be demonstrated through documented analysis of privacy risks and mitigations. These privacy processes should be internationally-based and interoperable, to enable the free flow of data needed to maximize innovation.

- *Intelligibility.* When AI is used to help make decisions that impact people's lives, it is important for individuals to understand how those decisions are made. But achieving useful explanations of the behavior of AI systems and their components—known as "intelligibility" or "explainability"—can be quite complex and highly dependent on a host of variables, precluding a "one-size-fits-all" approach. This is an area of intense, cutting-edge research and both industry and academia are actively exploring emerging methods for enabling intelligibility, as well as the scenarios in which, and reasons why, intelligibility may be required. Promising technical approaches have begun to emerge in achieving intelligibility for both system components (such as data and individual models) and entire systems. NIST should support the development of such approaches and participate in research regarding whether the behavior of AI systems, or components of those systems, is understandable to humans.

- *Bias and Fairness.* The fairness of AI systems is one of the key concerns facing society, as AI plays an increasingly important role in our daily lives. Fairness of AI systems must be a first-order priority, much like security and privacy. While it is not possible to "guarantee" fairness or "de-bias" AI systems, NIST should examine and support specific mechanisms that may be used to prioritize fairness in AI. AI systems can be unfair or biased for a variety of reasons, including societal biases that are reflected in the data used to train the system or biases reflected in the decisions made by teams (explicitly or implicitly) during the AI development and deployment lifecycle. In other cases, AI systems behave unfairly not because of societal biases, but because of characteristics of the data (e.g., too few data points about some group of people) or characteristics of the systems themselves. It can be difficult to distinguish between these reasons, which are not mutually exclusive. Microsoft recommends that NIST support ongoing research on and development of processes and tools for detecting and mitigating unfairness and bias. Rigid requirements mandating the use of particular technical tools or metrics are unlikely to be effective across the many contexts in which AI is used. Instead, Microsoft encourages NIST to continue engaging in the development and documentation of organizational best practices for managing risks due to bias and unfairness in AI systems. NIST participation and contribution to the content of the Technical Report 24027 "Information technology -- Artificial Intelligence (AI) -- Bias in AI systems and AI aided decision making" in ISO/IEC JTC 1 SC 42 will facilitate multi-stakeholder dialogue and collaborations.

- *Data Quality.* Much has been written about how the training data used in developing and testing AI systems can influence the outcomes of the algorithms being developed, including how a lack of representative data may introduce bias into a system. Tools, mechanisms and broad best practices should also be introduced to help enable better characterization of the datasets, including the demographics and quality of the data collected. For example, Microsoft researchers are developing "datasheets for datasets"—a framework for documenting and

explaining the key characteristics of datasets, including their motivations, composition, how the data was collected and pre-processed, and any limitations that could result in unintended outcomes, such as known biases or violations of privacy restrictions.[56]

- *Data Sharing*. AI also relies on the ability to share data. Microsoft recommends that NIST identify further efforts that can support sharing of organizations' datasets, such as by establishing governance practices for responsible curation and sharing of datasets specifically intended for the training of models. Some existing standards address technical factors associated with data formats and schema, data interoperability, and data portability. NIST should collaborate with stakeholders to consider if additional standardization work is appropriate for foundational, technical or management standardization around data provenance and quality, common data taxonomies, data de-identification requirements, and other qualifications of shared data sets. Pre-existing standards that can be adapted for use in the AI context include technical standards from the International Standards Organization[57] and industry guidelines. In addition, NIST should consider its own role in providing reference data sets that could allow others to test or verify their models against the reference data set, such as to test reliability or bias.

Microsoft suggests that NIST work within ISO/IEC JTC 1 to obtain broad input on the potential gaps in existing standardization efforts in these and other areas. After receiving such input, NIST should consider how existing effort may be complemented by additional work by international standards bodies to support the adoption and implementation of reliable AI.

B. Increase NIST Involvement In Development of Tools and Best Practices, Including Supporting Use of AI to Deliver Government Services

Microsoft also recommends that NIST become more involved in the development of tools and other mechanisms to develop and deliver reliable AI. In particular, NIST can enable adoption of reliable AI by working with U.S. agencies to identify tools and other mechanisms needed to support the adoption of AI by the agency. Such efforts would promote the use of trustworthy AI systems in both the public and private sectors.

Some Federal agencies are already using AI to better serve their constituents and to deliver on their missions. For example, the Department of Homeland Security is piloting an anomalous analytics

[56] Timnit Gebru, *et al.*, *Datasheets for Datasets* (last revised Apr. 14, 2019), https://arxiv.org/abs/1803.09010.

[57] *See, e.g.*, International Organization for Standardization*, ISO/IEC DIS 22624: Taxonomy Based Data Handling for Cloud Services* (2019), https://www.iso.org/standard/73614.html; International Organization for Standardization, *ISO/IEC 20889: Privacy Enhancing Data-Identification Terminology and Classification of Techniques* (2018), https://www.iso.org/standard/69373.html; International Organization for Standardization, *ISO/IEC 19944: Data Flow, Data Categories and Data Use* (2017), https://www.iso.org/standard/66674.html.

(continued…)

capability that uses AI to detect malicious activity across networks.[58] In addition, the Explainable AI program at the Defense Advanced Research Projects Agency ("DARPA") aims to create machine learning techniques that produce more explainable solutions while maintaining high performance and appropriate levels of trust in the system.[59] More broadly, policymakers have recognized the ability of AI to improve delivery of government services. At the federal level, lawmakers last month introduced the Artificial Intelligence in Government Act, which would require federal agencies to develop governance plans for AI and establish best practices for identifying and mitigating bias and other unintended consequences.[60] At the state level, lawmakers in California are weighing a bill that would stablish a commission to seek input on how AI could be used to improve state services and propose ways to incorporate AI demonstration projects into existing government services.[61] Microsoft recommends that NIST engage with agencies to identify additional tools and best practices that could support similar effort to use AI in delivering government services.

C. Broader Efforts to Support Ongoing AI Standardization Work

Microsoft encourages NIST to work hand-in-hand with U.S. industry and the international community to finalize the foundational standards developed by ISO/IEC JTC 1 SC 42 and other working groups. Those standards should then be adopted as American National Standards and used as the baseline for policies relating to the U.S. government's use of AI as well as any industrial policies or regulations that address the use of artificial intelligence in the U.S. marketplace.

Looking forward, NIST can also assist in advancing AI assurance efforts as the technology matures. Those efforts could focus on developing processes to ensure the reliability, safety, and robustness of AI. Tools that would support these outcomes may include provision of unique data sets, reliability testing models, and a catalog of reliability and fairness tools. Moreover, NIST can support key high-value scenarios with research, tools, and practices. This could include establishing government use cases and enabling subject matter experts to develop new tools. Such efforts may be most effective once a foundational set of internationally-developed, consensus-driven AI standards are in place.

NIST's existing Cybersecurity Framework and its work on the forthcoming Privacy Framework also address two areas of critical importance to AI technologies. As noted above in relation to Cybersecurity, in future revisions of these references, Microsoft recommends NIST to undertake a process to involve stakeholders in considerations about future updates to address potential issues arising from AI.

VI. Conclusion

[58] *See* Report to the President on Federal IT Modernization, Appendix C: Challenges to Implementing Federal-Wide Perimeter-Based Security (2017), https://itmodernization.cio.gov/report/appendices/challenges-to-perimeter-security/

[59] *See* Defense Advanced Research Projects Agency, Explainable Artificial Intelligence, https://www.darpa.mil/program/explainable-artificial-intelligence.

[60] *See* AI in Government Act of 2019, H.R. 2575 (introduced May 8, 2019), https://www.congress.gov/116/bills/hr2575/BILLS-116hr2575ih.pdf.

[61] *See* Artificial Intelligence in State Government Services Commission, A.B. 976 (introduced February 21, 2019), https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201920200AB976.

Microsoft appreciates NIST's commitment to undertaking a thorough examination of how standards and related tools can be used to support reliable, robust, and trustworthy systems that use AI technologies. NIST is in a unique position to assist in these efforts. Microsoft encourages the U.S. Government broadly to demonstrate its commitment to high-level ethical and moral principles needed to develop and use trustworthy AI. Moreover, and as a complement to those efforts, Microsoft encourages NIST to identify gaps in existing AI standardization efforts and lead efforts to address any such gaps within the process established by ISO/IEC JTC 1. These efforts will both protect and enable all organizations as they expand use of AI systems by fostering the development of trustworthy AI.

Respectfully submitted,

Jason Matusow
General Manager, Corporate Standards Group
Microsoft Corporation

June 7, 2019