July 19, 2019

Intel Comments on the NIST draft "Plan for Federal Engagement in Developing AI Technical Standards and Related Tools"

Intel Corporation appreciates the opportunity to provide comments on the NIST draft "Plan for Federal Engagement in Developing AI Technical Standards and Related Tools".

Intel is a world leader in computing and technology innovation. The company designs and builds essential technologies that serve as the foundation for consumer products, commercial systems and infrastructure equipment. Intel also invests in the development and adoption of global standards which have enabled advancements and interoperability of products and systems worldwide.

Introduction

Intel endorses NIST draft plan's focus on federal engagement in consensus standards, recognizing the importance of openness, global relevancy and a diversity of organizations to serve U.S. interests. Intel also commends the draft plan's comprehensive approach including: recognizing the importance of flexibility due to rapid advancements in AI, investment in early or pre-standardization stage areas, application of existing technical standards (especially general IT standards) wherever possible, and roles of horizontal and vertical sector-application standards. Intel also supports identifying prioritization aspects for standards and tools development, promoting areas of important research, emphasizing need for ongoing cooperation (including public-private-partnerships), and continued review of standards engagement plans with the private sector.

Intel supports NIST's direction for the plan to follow U.S. policies which emphasize voluntary, private sector-led consensus standardization (OMB Circular A-119 "Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities") and promote innovation and competition. Intel further supports the elaboration (section 2A) on the characteristics of consensus-based standards efforts that are important for Federal engagement prioritization.

We respectfully offer additional consideration areas for enhancing the draft plan, including general recommendations for each area. A summary of specific recommendations is included in the table at the end.

Intel welcomes the opportunity to further discuss these inputs as NIST develops the plan.

Standardization in Trustworthiness Space

We commend the emphasis on trustworthiness as a core requirement of AI systems. We expect the focus of the plan to “promote focused research to advance and accelerate broader exploration and understanding of how aspects of trustworthiness can be practically incorporated within standards and standards-related tools” will lead to better application of existing trustworthiness related standards in AI systems, in addition to new gap-filling standards specific to AI systems.

Automating the standardization landscape scan and gap analysis

Summary: It is important to evaluate a broad selection of potentially applicable standards, specifications, and best practices for an optimal landscape scan. The body of work relevant to AI is so large that we suggest some level of automation for a comprehensive and objective analysis.

The Federal plan mentions the following as part of the Plan:

- Conduct a landscape scan and gap analysis to identify standards and related tools that exist or need to be developed.
    - If appropriate standards exist, use them.
    - If appropriate standards do not exist, engage in their development (with specific activities associated with this engagement)

There are thousands of potentially relevant standards, specifications, best practices, and other documents in the broad AI area. While many lists of AI standards have been constructed, compiling these lists by hand, even in a representative group of experts, is likely not to be comprehensive and miss some potentially relevant subfields. Furthermore, due to the volume of information, it may be difficult to study potentially applicable documents in depth and identify commonalities or differences between standards or gaps in standardization reliably.

Statistical text processing tools have been developed to support automatic and semi-automatic categorization and to estimate the degree of relevancy of documents in a document set. When further analysis is needed, ontologies may be created automatically and semi-automatically to formalize knowledge representations for a topical area. With ontologies, reasoning and other techniques can be employed to detect affinities and conflicts and assess gaps. Standards and technical specifications are typically highly structured documents, lending themselves well to automated pre-processing, and making it easier to detect potentially pertinent common building blocks in standards and specifications that may not appear relevant based on quick scan of the title and abstract or scope.

While assessment by experts is still necessary to make sense of the results of the automated pre-processing and translate these results into a strategy, automating the initial stages of the landscape scan could greatly reduce the effort needed and provide greater coverage of the field and its subfields. It can also add a level of precision to the gap analysis work. Additionally, it may allow experts to focus their effort in areas that require subject expertise, rather than on automatable tasks.

General recommendation: consider tools to automated pre-processing of document sets and building blocks they contain to efficiently carry out a comprehensive landscape scan.

Benchmarking Aspects

We commend the plan's coverage of benchmark programs for evaluating the performance of AI systems. However, some specific areas of benchmarking activities as well as integration of different types of benchmarking need to be emphasized.

General Recommendations:

1. **Benchmarks should include "classical" AI and machine learning approaches, such as kernel methods and decision trees, in addition to deep learning (e.g., MLPerf).**
2. **Benchmarks and other forms of evaluation should consider end-to-end performance of the entire system, including the processing steps that are integral to the AI system (such as loading and pre-processing data, post-processing, storing results) but are not strictly part of the AI algorithm itself. In general, many real-world AI workloads are heterogeneous, and evaluation methods should reflect this.**
3. **Where possible, evaluation methods should include the total cost of owning and operating the system.**

Taken together, these steps will help increase the likelihood that benchmark results are consistent with real-world costs and performance and include the full diversity of approaches to AI systems.

We also offer a few comments on Appendix III. Each benchmark suite description in Appendix III can be expanded to align the content in the description. For example, each description can include information about the benchmark ownership and maintenance in general terms (e.g., whether it is created and maintained by a single entity or a consortium), its release date and last update, its broader scope, how it is used (if available), and levels of adoption, e.g., by referring to trends in the number of downloads, citations etc.

Privacy and Data Aspects

Privacy is a significant component of the elements that constitute trustworthiness, and it is especially important for Artificial Intelligence.

Privacy represents an intrinsic and foundational value for our modern society. To achieve trustworthiness of AI, and a trusted adoption and deployment of the technology, privacy is one of the aspects to be considered early in the design of AI technologies.

The creation of data standards – as suggested in the draft plan – would enhance access to data and the possibility to train datasets for different ML applications. For this purpose, categorizing data would be helpful to define individuals' privacy risks linked to data processing, but such a taxonomy would be valuable only if aligned with international standards and best practices and built by the international community. Differences in data classification are an obstacle to this outcome. Harmonization can help overcome these differences and incentivize data sharing where it is important. For example, in some countries, use of anonymization techniques are not sufficient to qualify the information as non-personal (e.g. anonymized medical data would still be considered highly sensitive personal identifiable information). International standardization can provide an avenue to achieve consensus on best practices and develop technical approaches to classification.

International alignment in the privacy space that will affect AI trustworthiness would be achievable only if consensus could be built around common principles rather than quantifiable metrics. Privacy risks are different depending on the use case of technology, the type of data involved, the societal and cultural context and many other factors that would be difficult to measure and translate into specific measurable items to test AI applications against. Additionally, it is necessary to work on aligning risk management vocabulary and metrics for the different elements of trustworthiness (privacy, security, safety etc.). Addressing the combination of those risks still presents a challenge and requires new approaches to risk composition that need to be explored and developed by research.

Research is also needed in the area of privacy preserving machine learning, including homomorphic encryption, federated machine learning, and relevant aspects of differential privacy. In cases where initial standardization activities have started, such as homomorphic encryption (e.g., ISO/IEC JTC 1) or federated machine learning (e.g., IEEE), these efforts should be studied and taken into consideration.

General Recommendations:

- **Seek international alignment in data taxonomy to foster interoperability and data sharing;**
- **Avoid creation of metrics for privacy risks that would be hard to combine with other trustworthiness elements like security and safety;**
- **Rely on existing Privacy Principles.**

Societal Considerations

Successful AI systems will not only achieve technical goals and solve functional problems, but will inspire transparent, accountable, and trusted collaborations between technologies and the people whose interests they serve. To this end, consensus-based ethical principles are critical to defining guardrails for AI system development and deployment, and for mitigating risks to persons affected by AI applications.

Although strong commonalities currently exist between ethical principles proposed by a broad base of experts (e.g., OECD, IEEE, WEF), tighter cross-sector alignment across and within ethical themes such as privacy, bias, transparency, and equity will be critical to addressing two current challenge areas: 1) the development of connections between AI ethical principles and standards; and 2) the development of tools to define and assess the type, likelihood, and magnitude of ethical risks posed by AI applications to persons in particular social contexts. These issues should be first addressed by relevant research.

Working toward the first challenge can assist practitioners tasked with integrating theory and practice throughout product development lifecycles[1]. Ethical standards can guide efforts towards, for example, building ethical checkpoints into existing processes, similarly to currently common product privacy reviews, as well as developing internal tools for training, assessment, oversight, and auditing, and providing guidance to suppliers and business customers.

The second challenge area, the development of AI ethics-oriented risk assessment tools, will be critical to predicting and monitoring gaps between anticipated and achieved deployment outcomes. The scarcity of tools to assess risk in particular social contexts is currently slowing efforts to deploy trustworthy AI. The Partnership on AI notes[2], for example, that effective use of existing risk assessment tools in the context of criminal justice is hampered by three main factors:

1. Concerns about the validity, accuracy, and bias in the tools themselves;
2. Issues with the interface between the tools and the humans who interact with them; and
3. Questions of governance, transparency, and accountability.

General recommendations:

1. **Work toward developing consensus on AI ethics principles, cross-sectors and within ethics related themes;**
2. **Map consensus-based AI ethical principles to ethical standards with consideration for product development lifecycles;**
3. **Research and develop valid, accurate, and unbiased AI ethics-oriented risk assessment tools**

[1] see e.g., www.partnershiponai.org/ai-ethics-requires-a-bridge-between-theory-practice

[2] www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system

Summary of Specific Recommendations

Sec. #	Line	Rationale	Recommendation
1(E)	195-197	Ensure that use cases take into consideration the range of outcomes for affected individuals, which vary as a function of the social context in which the system will be deployed. Affected individuals include not only targeted users, but also bystanders and other members of communities who are spatially or temporally located near a target.	Note that use cases must be accompanied not only by explicit information about the parameters of use, but also by the practical implications of such uses for persons who may be affected by AI deployments.
1(E)	202-213	In modern ecosystems, data travel across national boundaries; therefore, the international nature of data movements needs to be considered as a key requirement in standardization	Stress the need to align data standards with international standards and regulations, due to their key characteristics.
1(E)	212-213	Identifying and methodically documenting unintended consequences of AI deployments is crucial for adjusting parameters in subsequent iterations to mitigate harm to affected persons.	Include as aims of creating tools for accountability and auditing, not only "…to provide a record of events such as their implementation, testing, and completion," but also to assess and document gaps between predicted and achieved outcomes.
1(F)	259-263	The development and adoption of a common set of ethical principles can incentivize research collaboration and reduce friction toward standardization efforts.	In this area, standards flow from principles, and thus a first step toward standardization will be reaching broad consensus on a core set of AI ethics principles.
1(F)	257-258; 267-268	A current challenge area is mapping connections between AI ethical principles and standards.	Emphasize the value of researching approaches to a tight coupling between principles and standards.
1(F)	269	The standardization of ethical principles will require close attention not only to the degree of risk posed to humans, but also to the nature and likelihood and consequences of risk within areas, such as privacy, bias, and transparency.	The degree to which ethical considerations might be incorporated into standards should be tightly connected to the type, likelihood, degree, and consequence of risk to humans.
1(F)	270-271	Privacy requirements and implementations depend on context and the nature of data	Take into consideration the fact that privacy risks are different depending on the use case of

Sec. #	Line	Rationale	Recommendation
		sets, which may not always contain Personally Identifiable Information.	technology, the type of data involved, the societal and cultural context and many other factors.
1(F)	272-275	Existing ethics risk assessment tools are hampered by concerns about the validity, accuracy, and bias in the tools themselves; issues with the interface between the tools and the humans who interact with them; and questions of governance, transparency, and accountability.	Note the value of researching and developing valid, accurate, and unbiased AI ethics-oriented risk assessment tool.
2(A)	365-366	To make research and operational progress, it will be critical to prioritize such topics as transparency, bias and privacy.	Include "maximizing transparency" in the bullet that also names "identifying and minimizing bias."
2(C)	397-98	Fully manual landscape scans may not be efficient due to the size of the space and is likely to leave out significant subareas.	Make use of automation techniques (modern statistical methods) to assess relevancy of potentially related standards and specifications and to conduct primary analysis of relevant documents.
3	464-467	Maintaining a "flexible posture" is critical for the U.S. to adapt to the rapid pace of AI technology advancements, developing standards, and evolving requirements based on latest use cases, applications, and consensus around human-centered implications of AI. The best-practice policy approaches for sustaining flexibility define the high-level objectives and requirements (for the technical policy, regulation or procurement) and rely on the use of voluntary standards for the implementation. Existing standards which meet the objectives and requirements can be referenced as examples and use of equivalent standards are recognized or allowed.	Replace the sentence: "Maintain a flexible posture in specifying AI standards that are referenced in regulatory or procurement actions" With the following edited version: "Maintain a flexible posture in identifying suitable AI voluntary standards for referencing in regulatory or procurement actions, including allowance for equivalent standards that meet the objectives and requirements of the actions. "
3	502-504	Due to the nature of AI environments, the benchmarking process needs to include a combination of approaches, from technical to economic considerations,	Combine technical assessment, end-to-end performance evaluation, and total cost of ownership as necessary

Sec. #	Line	Rationale	Recommendation
			components of broadly applicable benchmarking techniques.
4	510-512	AI makes use of the global infrastructure, and global collaboration is necessary, in part via international standardization, to promote greater harmonization of AI standards which meet U.S. needs.	Remove “like-minded countries” and replace with “other countries” as this limitation may be an obstacle to collaboration in international standards bodies and with international counterparts on the variety of subjects related to AI.
Apx III	760-793	Benchmarking tools listed in Appendix III are very diverse, but their differences and the fields of applicability are not clear from the information in the Appendix.	Each benchmark suite description in Appendix III should be expanded to provide further information on the nature, positioning, and applicability of the tool.