July 1, 2019

Comment Template for Draft Plan for Federal Engagement in Developing Technical Standards and Related Tools for AI Technologies

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
1	Booz Allen Hamilton	Editorial Major	Section 3.3; Page 17	Crucial to ensuring the integrity and trustworthiness of AI systems is ensuring a level of consistency in the quality and provenance of training data utilized to build AI models and systems. As it pertains to challenges related to robustness, developing standards for data provenance now will provide additional protection against the threat of data poisoning in the future. In conjunction with federal agencies, NIST should coordinate and facilitate improvements in data sharing to support building better AI systems. While Section 3.3 contains language around how the government and private sector can explore non-traditional collaborative models such as open data initiatives, there is an opportunity to add more specificity and focus for different AI applications, such as identifying initial priority areas for focused efforts around dataset discoverability, identification, and shareability. Issues of data ownership and licensing are significant within the context of the federal government systems where many open source licenses prohibit the use of data in contexts where the output will not be made public.	Add language to direct the supplementing of data discoverability / shareability initiatives with the ability to search data by license
2	Booz Allen Hamilton	Editorial Major	Section 3.3; Page 17	While Section 3.3 contains language around how the government and private sector can explore non-traditional collaborative models such as open data initiatives, there is an opportunity to add more specificity and focus for different AI applications, such as identifying initial priority areas for focused efforts around dataset discoverability, identification, and shareability. Issues of data ownership and licensing are significant within the context of the federal government systems where many open source licenses prohibit the use of data in contexts where the output will not be made public. One solution to this challenge could be to supplement data discoverability / shareability initiatives with the ability to search data by license to make this information more readily available. As NIST identifies ways to increase data discoverability and access to Federal government data, it should begin by identifying, organizing, and sharing datasets that align to priority areas and missions with a heavy AI focus. For example, given the Department of Defense and Joint AI Center's organization around the National Mission Initiatives (NMIs) primed for AI development, NIST could begin facilitating the development of datasets and appropriate standards for the curation of datasets that align to the NMIs around humanitarian assistance and disaster relief (HADR) and predictive maintenance. Additionally,	Add language to clarify NIST's role in identifying and prioritizing the curation and development of datasets, standards, and sharing practices that align to missions or verticals with heavy AI focus such as humanitarian assistance and disaster relief or predictive maintenance

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
				with the interest and focus on building AI tools for healthcare or fraud, waste, and abuse applications, NIST should work with applicable agencies to identify, prioritize, and develop standards that are appropriate for those specific applications. This approach will ensure that the standards developed are appropriate for specific verticals while informing the evolution of broader horizontal standards.	
3	Booz Allen Hamilton	Editorial Major	Section 3.2, Page 17	Much of the language in the Draft Plan centers around elements that support the research and development of "trustworthy" AI. Trustworthiness standards, as defined in the draft plan, include accuracy, explainability, resilience, safety, security, and reliability; the ability to measure, track, and monitor such features is contingent on the ability to build auditable and explainable AI. Rather than discussing broadly how to conduct research or focus industry engagement on these specific areas, NIST should examine the intersection of these features and recognize that to achieve trustworthy AI, AI systems must be transparent and explainable. This should be tackled from the bottom-up; NIST should enact standards that require auditability be a feature that is built into AI systems, which can then elevate and inform the process for creating explainable and trustworthy AI. The level of trustworthiness or explainability should be tailored to the intended use case or scenario.	Add language to enact standards that require auditability be a feature that is built into AI systems, which can then elevate and inform the process for creating explainable and trustworthy AI. The level of trustworthiness or explainability should be tailored to the intended use case or scenario.
4	Booz Allen Hamilton	Editorial Major	Section 3.3; Page 17	The opacity of AI systems is one of the greatest barriers to trustworthy AI. Despite heavy research investment in AI over the last decade, in many cases, models are still black boxes, meaning that it can be impossible to parse or understand why the model reached a specific conclusion or recommendation. Models lack transparency and explainability, but the first step to ensuring that they are built into systems is to ensure that AI models are built to be auditable, meaning it is possible to have insight into the training data, model information, and other system information that inform the end decision, recommendation, or output. Auditability is key to understanding how and why AI systems arrive at conclusions or recommendations, and can then inform the appropriate explanation for the recipient of the information.	Add language to include the role of NIST to support and expand public-private partnerships to understand how to best tackle the challenge of explainability across broad industry verticals, and use the insights derived to inform horizontal standards.

Comment Template for Draft Plan for Federal Engagement in Developing Technical Standards and Related Tools for AI Technologies

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
5	Booz Allen Hamilton	Editorial Major	Section 3.3; Page 17	Section 1E of the Draft Plan calls out the need for tools for accountability and auditing to examine system output, traceability, or a record of events, but NIST has the opportunity to establish standards that influence the build of systems themselves that are auditable, rather than separate tools to perform this function. If AI systems are built to be auditable from the start, these additional tools for accountability and auditing will only be needed for independent verification and validation, which will ensure another layer of protection and improved performance. Building auditability into systems as a part of the fundamental design ensures that there is transparency into how systems make decisions. Auditable AI systems should contain key metrics and information around system build and performance, which can be sourced by engaging with the appropriate agency and industry groups to identify the most salient features for agency/industry verticals. Like version history in software releases, AI audit trails should at a minimum contain information such as model version (source of the original model, the technique used to train it, performance metrics around accuracy, when it was last tuned), and data provenance (source of the training data). At an even more granular level, systems should be built so that this information can be accessed / tracked in real-time, so that it is possible to parse this information at the time of inference. Many of these standards with respect to model versioning provenance, audit trails, etc. are likely already part of organizations' internal software development practices and aren't necessarily shared publicly.	Add language to discuss NIST's role in identifying, validating, and communicating best practices through its relationships with the broad machine learning research community, as well as its own research
6	Booz Allen Hamilton	Editorial Major	Section 3.3; Page 17	By approaching the trustworthiness issue from the system-level, NIST can drive agencies and industry to common standards and metrics around how AI systems should be built, what information needs to be captured to establish performance evaluation metrics and benchmarks, and ultimately, stronger risk management and mitigation procedures. The all or none nature of ecosystem ownership especially with respect to data access and the sensitivity of the models means that private organizations may not be highly incentivized to produce easily auditable standards for their models, but it is up to organizations such as NIST to drive progress against goals around trustworthy AI.	Add language around how NIST should work with industry to discover and evaluate value creation from making trustworthy AI. If organizations can recognize additional value from market-driven acceptance of AI because AI trustworthiness is perceived as having more utility, then those organizations will be more apt to pursue it on their own, which will lead to more and better approaches for tackling the challenge.

Comment Template for Draft Plan for Federal Engagement in Developing Technical Standards and Related Tools for AI Technologies

COMMENT #	NAME OF COMMENTER	TYPE i.e., Editorial Minor Major	LINE # PAGE etc.	RATIONALE for CHANGE	PROPOSED CHANGE (specific replacement text, figure, etc. is required)
7	Booz Allen Hamilton	Editorial Major	Section 3.3; Page 17	The Draft Plan is lacking with respect to addressing the unique challenges associated with AI security. Due to the lack of standards and governance in how they are built, AI systems are susceptible to adversarial threats, which can seriously threaten the integrity of the data and models that inform system output. If AI systems aren't robust and secure against adversarial attacks, they cannot be trustworthy, regardless of any efforts taken to build features that enable auditability or explainability. Both types of adversarial attacks (referenced in the text box) can have serious consequences when deployed, because the information / recommendation provided will be informed by false data, or a faulty model. Section 2A of the Draft Plan highlights important standards characteristics that warrant Federal government consideration, and while several of the areas address components of the adversarial AI challenge, none specifically speak to the importance of building secure, robust systems that can withstand adversarial attacks.	NIST should add language around engaging industry to conduct research and identify ways to build proactive defensive measures against adversarial attacks into AI systems.
8	Booz Allen Hamilton	Editorial Major	Section 3.1; Page 16	Pursuant to the recommendations in Section 3.1, crucial to bolstering AI standards-related knowledge, leadership, and coordination among agencies is establishing common governance standards for AI tools and solutions. As part of the National Science and Technology Council (NSTC) Machine Learning / Artificial Intelligence (ML/AI) Subcommittee's efforts to gather and share AI standards-related needs, it should examine and identify needs for common governance controls and practices. The current process for developing and deploying AI models is largely piecemeal, with bespoke models developed and deployed for specific data or problem sets. Rather than considering how to best develop and deploy AI that can be operationalized at enterprise scale, groups operate independently of one another to create AI models and tools that work for their problems. This approach, defined by a lack of coordination and oversight, exposes organizations to serious risk, because there is no way to establish common governance controls and procedures. By establishing common standards around governance that can be adapted for vertical standards to support individual agency or industries' needs, NIST can ensure the appropriate level of governance is adopted commensurate to the deployment, thereby reducing risk.	NIST should add language to Section 3.1 to define its role in oversight of the establishment of common AI governance controls.