Date: July 11, 2019

To: The National Institute of Standards and Technology

From: Dr. Michael Stumborg, Dr. Christine Hughes, Center for Naval Analyses

Subj: A Plan for Federal Engagement in AI Standards (draft for public review, 2 July 2019)

Encl: (1) FOIA Exemption Definitions
(2) Security Classification Definitions
(3) Narrow and General AI Definitions

Disclaimer

Researchers from the Center for Naval Analyses, the Department of the Navy's Federally Funded Research and Development Center, hereby submit this response to the subject NIST Plan. The information herein is the opinion of the authors only. It does not necessarily represent the opinion of the Department of Defense, the Department of the Navy, or the Center for Naval Analyses.

Following our submission to the original NIST RFI on AI standards, we were extremely pleased to see the inclusion of data accessibility and data quality standards in the draft plan. We respectfully submit the recommendations below, in hopes of improving what is already a very fine plan for accelerating U.S. leadership in AI. We thank you for the opportunity to contribute.

Recommendations Regarding Data Accessibility Standards

We recommend against language that might conflate data visibility with data accessibility. We also recommend that data accessibility standards not be conflated with data quality standards. They should be treated as separate and distinct for three reasons:

1. Data access is required, even for poor quality data. Data consumers (to include AI system developers) are already accustomed to having to improve and validate the quality of data before using it. Facilitating data accessibility via data accessibility standards will go a long way toward accelerating AI adoption and leadership, because its quality can be improved by making it accessible to users with a vested interest in data quality.

2. To advance AI transparency (and thus, public trust and acceptance) data must be made accessible not only to AI system developers, but also to (authorized) external/third party oversight bodies *for the sole purpose of evaluating the data quality*.

3. Data accessibility standards are more easily achieved. Many of the data accessibility attributes (listed in lines 216-21) that could become data accessibility standards are already codified in law. For example, the FOIA exemptions and security classification levels (see Enclosures 1 and 2) are settled law. These terms are not subject to the valuable, but arduous consensus-building process that data quality standards will have

to undergo. Furthermore (and unlike data quality standards) data accessibility standards are less likely to be specific to particular stakeholder communities. Separating data accessibility standards from data quality standards, and developing data accessibility standards first allows NIST to achieve a "quick win" that will create further momentum for the remaining AI standards development (not just for data quality standards). Early establishment of a set of AI standards can also alert, and bring to the discussion, additional stakeholders whose input will greatly improve the development of the remaining AI standards.

With these observations in mind, we recommend the following changes to the plan:

Footnote 7: Change "big data analytics; data exchange; data quality; and data privacy." to "big data analytics; data exchange; data quality; data accessibility; and data privacy."

Line 214-5: Change "more visible and more usable" to "more visible, accessible and usable"

Line 463: Change "data access and quality" to "data accessibility and data quality."

Line 500: Change "access" to "accessibility"

Recommendations Regarding Definitions

We note from our work with non-technical senior leaders who are expected to develop and deploy AI systems, considerable confusion regarding the risk associated with AI. There is a pervasive misunderstanding of the difference between General, and Narrow AI. The vast majority of recent AI developments are limited to Machine Learning (ML), a form of Narrow AI. Machine Learning is engineering fact, but many decision makers base their risk calculus on the science fiction of General AI. We include in Enclosure 3, useful definitions for Narrow and General AI from the Office of the Director of National Intelligence and recommend these definitions be included in Appendix I of the plan.

We note multiple instances in the plan, of "AI and ML." "AI/ML," and "ML/AI," which all imply that ML is different from AI, when in reality ML is one subset of (Narrow) AI. The plan already addresses the need for standard definitions for AI terminology, and it would be premature for the plan to solidify definitions now without stakeholder engagement. We do recommend though, that the plan prioritize the development of standard definitions for the terms AI and ML first, and to ensure that the plan does not imply that these terms are already defined.

Recommendations Regarding Ethical Considerations and Risk

Lines 266-9 claim that consensus exists regarding the degree to which ethical considerations need to be incorporated into standards, implying that standards become less important where the risk to humans is of a lesser degree. We do not agree. Developers of AI systems with minimal, but non-zero risk to humans should still be expected to abide by ethical standards because failure to eliminate even minimal harm can and will be pointed to by detractors as evidence that AI is not to be trusted in applications where greater harm is possible.

Enclosure 1: FOIA Exemption Definitions

These definitions are available at: https://www.foia.gov/faq.html

- **Exemption 1:** Information that is classified to protect national security.
- **Exemption 2:** Information related solely to the internal personnel rules and practices of an agency.
- **Exemption 3:** Information that is prohibited from disclosure by another federal law.
- **Exemption 4:** Trade secrets or commercial or financial information that is confidential or privileged.
- **Exemption 5:** Privileged communications within or between agencies, including those protected by the:
    1. Deliberative Process Privilege (provided the records were created less than 25 years before the date on which they were requested)
    2. Attorney-Work Product Privilege
    3. Attorney-Client Privilege
- **Exemption 6:** Information that, if disclosed, would invade another individual's personal privacy.
- **Exemption 7:** Information compiled for law enforcement purposes that:
    - 7(A). Could reasonably be expected to interfere with enforcement proceedings
    - 7(B). Would deprive a person of a right to a fair trial or an impartial adjudication
    - 7(C). Could reasonably be expected to constitute an unwarranted invasion of personal privacy
    - 7(D). Could reasonably be expected to disclose the identity of a confidential source
    - 7(E). Would disclose techniques and procedures for law enforcement investigations or prosecutions, or would disclose guidelines for law enforcement investigations or prosecutions if such disclosure could reasonably be expected to risk circumvention of the law
    - 7(F). Could reasonably be expected to endanger the life or physical safety of any individual
- **Exemption 8:** Information that concerns the supervision of financial institutions.
- **Exemption 9:** Geological information on wells.

Enclosure 2: Security Classification Definitions

The three authorized Federal Government classification levels are defined in U.S. Code Title 50, Chapter 44, Subchapter VI, Section 3161:

(1) "Top Secret" shall be applied to information, the unauthorized disclosure of which reasonably could be expected to cause exceptionally grave damage to the national security that the original classification authority is able to identify or describe.

(2) "Secret" shall be applied to information, the unauthorized disclosure of which reasonably could be expected to cause serious damage to the national security that the original classification authority is able to identify or describe.

(3) "Confidential" shall be applied to information, the unauthorized disclosure of which reasonably could be expected to cause damage to the national security that the original classification authority is able to identify or describe.

Enclosure 3: Narrow and General AI Definitions

These definitions are from the *AIM Initiative: A Strategy for Augmenting Intelligence Using Machines*, Office of the Director of National Intelligence. Available at https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf

Artificial Narrow Intelligence (ANI): Also known as "Narrow AI" or "weak" AI, this is an AI system that is specialized for a single purpose and cannot be generalized. All current applications are ANIs.

Artificial General Intelligence (AGI): Also known as "General AI" or "strong" AI, this is an AI system that can handle any human intellectual task—memory, learning, abstraction, and creativity. There are no AGI systems in existence, although building an AGI has been the goal of the field since it was founded in the 1950s.