May 31, 2019

To: *National Institute of Standards and Technology*

Subject: *RFI: Developing a Federal AI Standards Engagement Plan*

The Association for Manufacturing Technology (AMT) data science and manufacturing technology team members would like to provide this response regarding the AI Standards notice in the Federal Register. The perspective taken on AI standards is based on the data layer with consideration to broader application layers such as Big Data, cloud infrastructure, mobile technology, and the Industrial Internet of Things (IIoT)).

AMT is an advocate for the advancement of manufacturing technology solutions and productivity. The organization supports its members through Market Access and Industry Intelligence. The primary business objectives are to assist members in achieving increased sales and decreased costs. As such, AMT services include advocacy, access to its business intelligence portal, benchmarking survey management, custom research for gauging market share or industry forecasts, and manufacturing workforce development through its Smartforce initiative.

The MTConnect Institute is a not-for-profit subsidiary of AMT and established the MTConnect standard. AMT participates in the continued development and maintenance of this standard to encourage interoperability between differing systems, machinery, or devices. U.S. Manufacturing Technology Orders (USMTO), a program administered by AMT, tracks manufacturing technology order information for over 200 product categories with participation from both builders and distributors. The International Manufacturing Technology Show (IMTS), the premier manufacturing technology event in North America is sponsored and managed by AMT.

Adoption of Data Standards within AI: Integrated Sourcing, Robustness, Validity, Integrity

Central to AI, a disruptive technology is the quality of data. To capitalize on the economic and strategic value of AI solutions or products, it is prudent to incorporate integrated data sources during the design or development stages. Sources include both qualitative and quantitative information collected from databases, surveys, interviews, custom applications or other research. This task may help with addressing bias in the data thereby the bias in models from the onset given the increased diversification of the input.

In addition to data sourcing via integrated information, relationships that exist within the base dataset(s) need to be robust. The lineage and linkage of the base information should be holistically understood from both the business and technical domains. This task will help define assumptions being made during multiple stages of an AI project from data input to how modeling parameters are adjusted. Robust relationships in diversified data help capture how the output is influenced more efficiently.

Validity of data entails realistic and consistent representation of the information. This task assists in evaluating data reliability. It may shed light on variances observed in results that are measured or modeled. Creating mechanisms to assess data validity such as scores, rates, using validation datasets, or hypothesis testing processes should be addressed in an AI standard.

Models, over time may need to be configured with greater orders of magnitude (e.g., not solely historical data) to improve on quality and mitigate diminishing returns. The opportunities for increasing model performance by utilizing big data, third party (external) data, and training data include the associated risks of compromised data integrity. Iterative methods to test and identify corrupt data including such data generated by human error should also be defined within an AI standard.

Adoption of Pattern Detection Standards within AI: Transparency, Measures

Machine learning is a branch of AI where patterns are discovered within existing data to identify potential behavior on unseen (future) data. Documenting modeling assumptions including how characteristics of the data, such as noise or outliers are handled must be highly transparent. Instantiating a baseline model to compare against further iterations (e.g., model calibration), anomalies, or other variations (e.g., differing algorithm) will allow for proactive model management and reduced AI risk. These tasks will help prevent forming inaccurate conclusions about behavior or other patterns as well.

The methods by which information is qualified must be clearly stated. Examples include calculated or derived measurements, identifying how bias is removed from a measurement where applicable, and creating indexes or groups. Since these transformations often affect the classification of data fed into a model, the results may be unfavorably impacted. Sensitivity and stress testing processes may indicate the magnitude of this impact and should be incorporated within standard best practices.

Adoption of Foundation Standards within AI: Guides, Profiles

The Cross Industry Standard Process for Data Mining (CRISP-DM) originally developed in 1996 continues to be an extensively used analytics process model. It could be referenced within an AI standard because of its significance to advanced analytics. The nature of CRISP-DM, like data exploration itself is cyclical. Critical dependencies between the stages are denoted in this methodology.

When beginning to investigate data, a hypothesis guide or template can be a useful tool to support business objectives. Steps in this template include:

1. Stating the hypothesis
2. Identifying the iteration
3. Stating the business value of the hypothesis
4. Identifying the metrics against which the hypothesis will be measured (its progress & success)
5. Identifying the risks and impediments
6. Identifying decisions in support of the hypothesis
7. Identifying the predicted outcomes that support the decisions

The value of this tool is in yielding improved collaboration between business teams and AI initiatives.

Creating inclusive data profiles or maps which serve as the building blocks of AI projects should be addressed within an AI standard. This entails three key areas: Data Discovery, Data Architecture, and Data Models. Data Discovery encompasses the structure and composition of information. Data Architecture covers where the information resides, how it is categorized, and its associated input and output flows to other areas within an organization's data landscape. Data Models include the linkages, cardinality along with relationships that the information may have to other entities both internal or external to an organization.

McKinsey Global Institute McKinsey & Company, *Notes from the AI Frontier Modeling the Impact of AI on the World Economy*, https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx, September 2018

AMT, *Data Science Explained*, https://datametrologyworld.wordpress.com/2018/03/07/data-science-explained/

AMT, *Data Methodologies,* https://datametrologyworld.wordpress.com/2018/10/01/data-methodologies/

Data Science Central, *Dean of Big Data Driving the Snakes of Data Science Out of Ireland*, https://www.datasciencecentral.com/profiles/blogs/dean-of-big-data-driving-the-snakes-of-data-science-out-of

Data Science Central, *DSC Webinar Series: 4 Ways to Tackle Common Data Prep Issues,* https://www.datasciencecentral.com/video/dsc-webinar-series-4-ways-to-tackle-common-data-prep-issues