

August 8, 2019

The Honorable Mick Mulvaney  
Director  
The Office of Management and Budget  
725 17th Street, NW  
Washington, DC 20503

RE: Notice of Request for Information: Identifying Priority Access or Quality Improvements for Federal Data and Models for Artificial Intelligence Research and Development (R&D), and Testing

Dear Director Mulvaney:

On behalf of the physician and medical student members of the American Medical Association (AMA), I appreciate the opportunity to submit comments in response to the Notice of Request for Information: Identifying Priority Access or Quality Improvements for Federal Data and Models for Artificial Intelligence Research and Development (R&D), and Testing (RFI). The AMA applauds the strategic focus of the Administration to address the questions outlined in the RFI. In early 2017, the AMA's physician leadership councils prioritized the development of policy and expertise to advance augmented intelligence (AI) applications in health care. As data is needed to power AI innovation, particularly machine learning, the RFI raises important questions concerning access to and quality improvement needed for federal data and models.

### **Brief Background: AMA Health Care AI Policy**

The AMA strongly supports federal and state policies that promote AI applications in health care that advance the quadruple aim and equity. Specifically, AI systems with health care applications should enhance the patient experience of care and outcomes, improve population health, reduce overall costs for the health care system while increasing value, and support the professional satisfaction of physicians and the health care team. And, such systems should promote equity and guard against reinforcing structural inequality or new disparities. It is an AMA priority that there is appropriate professional and governmental oversight for safe, effective, and equitable use of and access to health care AI. Ultimately, the policies developed by the Administration should facilitate the design of high-quality and validated AI systems—particularly those that will be used in health care.

The AMA urges the Administration to adopt policies that will ensure AI applications are designed and evaluated in keeping with best practices in user-centered design, particularly for physicians and other members of the health care team when the AI systems are used in health care. In addition, the AMA strongly supports data policies that will support AI system transparency and reproducibility while also safeguarding patients' and other individuals' privacy interests and preserves the security and integrity of personal information. AMA policy also promotes education for patients, physicians, medical students, other health care professionals, and health administrators to promote greater understanding of the promise

and limitations of health care AI. We have enclosed the AMA's report and policy on health care AI adopted this past June to provide additional context and we welcome discussing these additional issues in greater detail.

### **Additional Accesses for Data and Models to Improve AI R&D and Testing**

To improve access to data and models, the federal government should focus on federal data sets being available through an application programming interface (API), read by an API, or easily imported into programming languages, such as Python. Thus, the federal government should promote the use of JSON and GeoJSON formatting. Additionally, in reviewing the datasets on data.gov, common formats like CSV, XML, or Excel should be prioritized over datasets being available in PDF, Word, text, or JPEG format.

The federal government should also promote providing the option to have federal data sets available for all years or nationwide rather than requiring downloading and merging 50 datasets for each year. While these individual datasets can be merged through programming, the process is tedious and introduces unnecessary errors. Relatedly, the federal government should also attempt to maintain the same column names for a field or attribute across multiple measurements or time periods. For example, in the Centers for Disease Control and Prevention's Social Vulnerability Index data set, the "proportion of the population below the poverty line" attribute was labeled G1V1R in 2000, E\_P\_POV in 2010, and EP\_POV in 2014.

Testing is a core aspect in the development, validation, and evolution of all technology. As with any technology, AI testing must ensure it meets the requirements that guided its design and development, responds correctly to inputs, performs its functions within an acceptable time, is sufficiently usable, and achieves the general result its stakeholders desire. Testing should also closely resemble operational production environments (e.g., a real-world medical facility), rather than synthetic simulations. To accelerate AI testing, the federal government should ensure federal datasets maintain their approximation of real-world care and practice settings. These efforts should be in coordination with the physician and medical informatics community.

### **Quality Improvements to Accessible Data and Models**

One of the most important characteristics to make data sets and models well suited for use by AI is requiring federal agencies to establish ethical governance, conscious design, and a learning culture regarding federal data. Establishing this type of federal data strategy, as laid out in the President's Management Agenda, may help increase the accuracy, precision, validity, and reliability of an AI's output, which could potentially increase the output's usability and trustworthiness (e.g., accounting for bias and fairness; transparency and explainability; and robustness, security, and safety).

Key aspects of a federal data strategy to improve the quality of the data include protecting privacy, ensuring sound data security practices, promoting transparency, and protecting integrity.

Health care information is one of the most personal types of information an individual can possess and generate. An individual's privacy should be honored unless waived by the individual in a meaningful way, de-identified, or in rare instances when strong countervailing interests in public health or safety justify invasions of individual privacy or breaches of confidentiality. Individual trust in the health care

The Honorable Mick Mulvaney

August 8, 2019

Page 3

system can only be assured when all entities—including the federal government—that maintain an individual's health information have an obligation to ensure the confidentiality of that information and when individuals have autonomy and control over decisions to disclose or retain their personal information. This type of control requires transparency from data holders to empower individual decision-making. Overall, individuals should have appropriate access to their data and physicians should have the tools and controls they need to be good stewards of this information, including their patient's medical record.

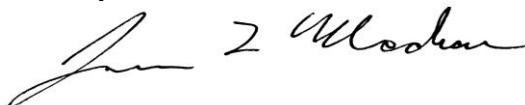
The federal government should also mitigate privacy risks as it relates to providing access to federal data. Privacy risks include re-identification of individuals through de-identified (or partially de-identified) data; misunderstanding or disregard of the scope of an individual's consent; individual perception of loss of their privacy leading to a change in their behavior, embarrassment, or stigma resulting from an unwanted disclosure of information or from fear of a potential unwanted disclosure; perceived and real risks of discrimination, including employment and access to or costs of insurance; and law enforcement accessing data repositories beyond their intended scope.

Ensuring security means emphasizing the need for security practices to stay up to date with current and emerging threats to protect data integrity and foster innovation and leverage new technologies to maintain protection. Promoting transparency includes supporting the transparency of data uses by stakeholders, organizations, and health care providers. Transparency includes how information is being shared, who is using the data, what data are being used for, and whether the data are being used for their consented purpose. Integrity of data should be ensured throughout its lifecycle including the collection, creation, analysis, use, storage, distribution, disclosure, or disposal of individual information.

The AMA believes that the promotion of efficient use of data access includes open access to appropriate machine-readable public data, development of a culture to share data with external partners, and explicit communication of allowable use with periodic review of informed consent. In providing open access to appropriate data, the AMA encourages disclosure of the characteristics of the datasets including the data sources, data collection, and data curation methods accompanied by an assessment of undisclosed and unintended biases resulting from the data-gathering process and any efforts made to mitigate these risks. Accounting for unintended bias in data sets is a central metric of data quality and a key to mitigating the risk of potentially furthering disparities.

The AMA appreciates the opportunity to comment and we welcome engaging further on this topic. If you have any questions, please contact Shannon Curtis, Assistant Director of Federal Affairs at 202-789-8510 or [shannon.curtis@ama-assn.org](mailto:shannon.curtis@ama-assn.org).

Sincerely,

A handwritten signature in black ink, appearing to read "James L. Madara". The signature is written in a cursive, flowing style.

James L. Madara, MD