AMERICAN PSYCHIATRIC ASSOCIATION

800 Maine Avenue, S.W.
Suite 900
Washington, D.C. 20024

August 9, 2019

Russell T. Vought
Acting Director
Office of Management and Budget
725 17th Street, NW
Washington, D.C., 20502

RE: Request for Information: Identifying Priority Access or Quality Improvements for Federal Data and Models for Artificial Intelligence Research and Development

Dear Acting Director Vought:

The American Psychiatric Association (APA), the national medical specialty society representing more than 38,500 psychiatrists who treat mental health disorders, including substance use disorders, appreciates the opportunity to submit feedback to the Office of Management and Budget (OMB) on its Request for Information: Identifying Priority Access or Quality Improvement for Federal Data and Models for Artificial Intelligence Research and Development. The APA is committed to ensuring that its members and the field of psychiatry are at the forefront of evidence-based care, innovation, and the continually-evolving digital healthcare landscape. The field of artificial intelligence within psychiatry holds exceptional promise in care delivery and patient outcomes and we would like to use this RFI not only to respond to the prompts therein, but also to highlight areas where AI can be used in patient care, as well as areas of potential concern.

Presently, the use of AI in healthcare is still fairly nascent. With respect to psychiatry, most AI-driven tools are embedded within health IT products more appropriately described as "augmented intelligence" rather than "artificial intelligence." In electronic health records (EHRs), the scope of this technology tends to encompass features such as electronic clinical decision support (eCDS). With mobile devices like smartphones, AI in apps could, for example, use the mobile device's sensors, the patient's tapping/typing activity, and the patient's behavioral history to warn of potentially triggering factors (e.g., risks for substance use).

While the APA is optimistic that the future of AI may improve patient care and lead to better outcomes, we are concerned that there are presently very few standards to which industry is being held in the development of AI in healthcare. For instance, standards around privacy, security, and confidentiality within health IT are currently in flux in the transition from ePHI under HIPAA to electronic health information (EHI), as it will soon be defined in the forthcoming 21st Century Cures Interoperability and

Information Blocking final rule. In the absence of clearly defined standards in AI, the APA is concerned that leveraging data for use in AI may compromise the privacy and security of patients' health information. **The APA recommends that future rulemaking should define such standards, using the expertise of agencies such as the Office of the National Coordinator for Health IT (ONC) and the Food and Drug Administration (FDA), which already has developed a pre-certification program for software as a medical device (SaMD).**

For example, algorithms embedded within AI may use sensitive information, such as a patient's genetic tests and substance use data to inform clinical decision support. Patients should ultimately own these data, be educated on when and how data is being released to developers for use in developing AI algorithms, and ultimately be empowered to provide consent to have their information de-identified using the Office of Civil Rights' standard 164.514(a)(b)(c), Expert Determination and Safe Harbor. Moreover, these precautions my not be enough to maintain the privacy of patient data. According to a recent article published in The New York Times[1], patient data may be easily reidentified. This is also concerning in the use of patient data by AI in healthcare. **The APA recommends that ONC and CMS review the adequacy of HIPAA standards overall in the deidentification process and specifically in the use of AI.**

The APA is also concerned with the generalizability of samples used in AI system research, development, and beta testing. If algorithms are developed using typical research samples, these algorithms may compute erroneous recommendations for treatment of individuals of a different sex, ethnicity or socioeconomic group. These algorithms may fail to help those with health comorbidities and unique vulnerabilities (e.g., developmental disorders, suicidal ideation, substance use disorders). Raw data in samples being used to develop current and future health IT, including AI, should be subject to verification, so that physicians can be assured that any algorithm recommending a specific course of treatment is validated across patient populations. **APA also recommends that AI research, development and testing be addressed through future rulemaking.**

What Federal data and models are you seeking to use that are available to the public with no use restrictions, but which have technical issues inhibiting data access? Specifically, what are the technical issues (e.g., is it too big to be downloaded, is it not optimally formatted)? What types of AI R&D and testing would be accelerated with increased access to this data? What research questions and applications are you trying to solve with AI that require specific types and/or quantities of Federal data an models, and how might the Federal Government reduce barriers to discovery and access?

There are nearly ~1,500 datasets under healthdata.gov. Many of these are discrete datasets. Some are part of a cluster of datasets among waves of longitudinal data. Some have very few cases and variables; some are more far-ranging. When considering how much healthcare data is available by the U.S. government, and using AI to leverage these data to further the "Quadruple Aim" in healthcare, these datasets should be merged and the variables recoded and transformed into various aggregated master

[1] https://www.nytimes.com/2019/07/23/health/data-privacy-protection.html. Accessed August 5, 2019

files, represented topically. Otherwise, the ability for various components of AI (machine learning, natural language processing, etc.) to make good use of the data will fall short of expectations.

For example, there are numerous datasets related to health IT adoption in data.gov that track trends of physician and hospital adoption as a result of the EHR incentive programs, starting with Meaningful Use and continuing into MIPS. As individual datasets, these are not particularly accessible or usable by researchers when employing AI in R&D. Examples of these datasets include:

- EHR Products Used for Meaningful Use Attestation
- Electronic Health Record Vendors Reported by Health Care Providers Participating in Federal EHR Incentive Programs
- Office-based Physician Health IT Adoption and Use
- State Health IT Privacy and Consent Laws and Policies, ONC Regional Extension Centers (REC) Key Performance Indicators (KPIs) by County

This is especially evident as these data files are not, to our knowledge, associated with any healthcare outcomes dataset. Having these datasets transformed, merged with each other, and merged with patient-level, hospital-level, and national datasets into a single dataset would aid in creating useful AI.

Having this information merged into a single unified dataset helps AI to detect trends in health IT adoption and implementation. A unified dataset could identify ways to mitigate the challenges that some physicians experience in integrating technology into their practice. And, when combined with health data, this unified dataset could reveal how adoption of health IT affects patient care and outcomes.

There are other examples how other datasets should be recoded, merged, and otherwise transformed to make AI useful. For instance, the "National Survey on Drug Use and Health" has multiple iterations and waves spanning years. To better leverage AI in tracking trends in drug use, these should be merged. The following datasets also have the potential for use in AI R&D, with respect to psychiatry:

- National Database for Clinical Trials Related to Mental Illness
- Healthcare Cost and Utilization Project Nationwide Readmissions Database
- DASH – Youth Risk Behavior Surveillance System: High School (which should be merged with the Middle School dataset)
- Drug Abuse Warning Network National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)—Wave 1 and Wave 2
- National Survey of Substance Abuse Treatment Services
- National Database for Clinical Trials Related to Mental Illness.

Other, similar epidemiological datasets could begin to help AI to detect trends in etiology and treatment/outcome patterns in mental health and substance use disorders. Future algorithms could be developed from detected trends for inclusion into electronic clinical decision support (ECDS) tools.

Connecting these seemingly disparate — but ultimately interconnected — datasets is imperative for useful AI. A useful AI can fill-in the gaps in our knowledge around the multidimensional aspects regarding the etiology and treatment of mental illness and substance abuse disorders. It can also inform treatment, the multidirectional trajectories ultimately experienced by many demographics of patients, including sex, ethnicity, socioeconomic status, and so on.

What Federal data and models are you seeking that are private and not at all available to the public? Describe the agency that has the data and what, if any, attempts you are aware that have been made to increase access to the data or model. What types of AI R&D and testing would be accelerated with increased access to this data?

De-identified, research-ready patient and hospital-level data could be accessed from the Department of Veterans Affairs, the Department of Defense, and Indian Health Service. These data — in combination with the epidemiological datasets above — could help to advance AI R&D for the reasons previously cited.

Finally, the APA also recommends that the CDC streamline the ability to request and receive identifiable data (with the appropriate research protections and permissions) in a timely and easy to use fashion from the National Death Index (NDI). This would permit record linkage with EHR data. Once linked, the data could be de-identified for subsequent analyses. Such data would be very useful for prospective identification of factors associated with suicide, with predictors of drug overdose deaths (including opioid related deaths), and associations between numerous other co-morbid mental and physical health conditions and patient mortality.

Thank you for this opportunity to respond to this RFI. If you have any questions, please do not hesitate to contact Nathan Tatro, Associate Director for Digital Health at ntatro@psych.org, or (202)-559-3680.

Sincerely,

Saul Levin, MD, MPA, FRCP-E
CEO and Medical Director