



Office of the Director
National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

March 15, 2024

Re: NOT-OD-24-037
Comments on Updated Strategic Plan for Data Sciences
Via: strategic-plan-rfi@od.nih.gov

To whom it may concern:

I am pleased to respond on behalf of the American Psychological Association (APA) to NIH's draft updated strategic plan on data science. The APA is the nation's largest scientific and professional nonprofit organization representing the discipline and profession of psychology, as well as over 157,000 members and affiliates who are clinicians, researchers, educators, consultants, and students in psychological science.

We commend NIH for developing this detailed plan and gathering comments from stakeholders. We especially acknowledge that the plan is written in inclusive language, so that psychological, social and behavioral science disciplines and scientists are reflected as communities to help shape and be shaped by the plan.

We also want to acknowledge the size and importance of this task. Taking advantage of the extraordinary opportunities that exist in the development of new technologies while also ensuring private data is protected, data can be shared to increase understanding and replication, and data can be harmonized to enhance its usefulness, are significant tasks separately. A plan that includes all those goals and more helps scientists think about the data they collect and use in more expansive ways.

Goal 1: Improve Capabilities to Sustain the NIH Policy for Data Management and Sharing

NIH priorities are well articulated; however, the implications for individual investigators are not completely clear. Investigators' ability to collect, analyze, report on and share data depend on many factors within and outside of NIH—available funding, peer review, training, for example. While many aspects of this plan are aimed at improving NIH's ability to steer and shape the tools that enable data management and sharing, we suggest that a summary of the plan be prepared for individual investigators with varying levels of resources who are working in an environment of increasing complexity.

The plan mentions the need to reach and involve lower-resource institutions, and this is an area that could be better developed. These institutions depend on NIH data just as others do. We also appreciate the acknowledgment of tribal ownership of data from Indigenous people.

Goal 2: Develop Programs to Enhance Human Derived Data for Research

We are pleased to see an emphasis on the many issues related to data collection, harmonization and reuse regarding Social Determinants of Health and Environmental Determinants of Health. We would also suggest you add “workplace” to the list of social determinants of health. There is a great deal of data about how the workplace significantly impacts both chronic diseases and mental health.

The strategic plan states, “To further advance NIH’s goal to bridge the gap between health care settings and applied and clinical research, NIH will strengthen the use of ontologies with vocabularies and terminologies (e.g., SNOMED, LOINC) and exchange standards such as FHIR®.” One NIH program that might be highlighted as a good example is the Office of Behavioral and Social Sciences Research’s [Ontologies](#) project. Before data can be harmonized, definitions need to be harmonized, and OBSSR is working to address this need in the behavioral and social sciences. Rigorous and reproducible behavioral science requires clear and consistent definitions of social and behavioral phenotypes, outcomes, and intervention components to allow effective communication between scientists and across scientific disciplines. This is an ideal project in which NIH’s influence and scope allow it to take the lead.

We appreciate and concur with the plan’s emphasis on training. Among other areas, the plan calls for cross-training between data scientists, clinical researchers, and nurses engaged in research; training on consent practices and ethical use of data that go beyond legal and regulatory requirements with special considerations for linked/merged data and data from underrepresented communities; and trainings on data sharing, management, transparency, provenance, and data quality for clinical research.

Goal 3: Provide New Opportunities in Software, Computational Methods, and Artificial Intelligence

As the draft plan states, “... technological innovations in AI and new capabilities to optimize large language models have generated considerable interest in the possibility of AI to recognize, summarize, translate, predict, and generate text and other content based on knowledge gained from massive datasets. Yet, challenges remain in creating transparent and explainable datasets and models. The need for ethical principles and frameworks, and connecting those principles/frameworks to practice, for developing and using AI in biomedical research remains an important priority and an unmet need.” The ways AI is transforming health care delivery will affect the available types and quality of research data. The data science plan rarely mentions mental health, so we want to provide some additional comments about needs in the mental health space especially in connection with AI technologies.

Health Care -- We are already seeing the deployment of AI-based tools to assist in the diagnosis and treatment of mental and behavioral health conditions. APA strongly supports the development of evidence-based guidelines and regulations regarding the use of AI technology in a healthcare setting, particularly in behavioral health care. AI technology must undergo rigorous design and development quality control processes to mitigate the biases to which it is susceptible. Ongoing research, some supported by NIH, is needed to determine the efficacy of the varied AI-based tools that currently are and will become available. Psychologists are uniquely positioned to contribute to these research efforts,

since our training equips us to study the impact of interactions between people and human-like AI-based tools on both the individual users and society more broadly. It must be noted that a lack of regulatory and reimbursement pathways may hinder the ability for these new technologies to scale. APA encourages continued collaboration between the FDA and CMS to develop regulatory and reimbursement pathways for the appropriate use of AI technology in mental health care. While clinical practice and reimbursement is not NIH's direct responsibility, we understand that the ability to develop and collect data on mental health interventions is directly affected by these issues and in some instances hindered by it.

Supporting Medical Innovation-- Artificial intelligence and innovative technologies have the potential to positively impact the health care system and the future of service delivery. For example, AI technologies may improve patient care by enhancing diagnostic precision, allowing for more individualized treatment, and improving engagement. Additionally, AI could foster the early detection of behavioral health concerns and enable the scaling of interventions to reach a much broader segment of the population than currently can access care. Technology aimed at identifying and treating social isolation, mental and physical health issues, and providing new tools from transportation to caregiving show promise in improving health.¹ Some of these tools provide a means for helping predict health risks or recommend personalized treatment plans, and others interface directly with patients in the form of therapeutic chatbots.²

Yet, the use of AI technology within mental health care raises several ethical considerations. First, the issue of bias: if the datasets used to train AI technologies are incomplete or reflect systemic biases, then the resulting AI technologies will also reflect those biases. Ultimately, this could exacerbate existing disparities and inequities. Ethics and diversity training for individuals involved in the development of new AI tools may help mitigate this effect. Second, given the sensitivity of mental health data, there needs to be considerable attention to data safeguarding to prevent public harm. Users of apps, chatbots, and other AI-based technology must be comprehensively informed about the privacy implications, including what personal data is being collected, how it is stored, who has access to it, and what happens to their data after they stop using the tool. Third, greater transparency is needed regarding the development and technology underlying AI-based tools (while protecting copyright and intellectual property), as well as rigorous testing of the AI-based tools to foster public trust, to establish efficacy, and to minimize unintended consequences. Finding a balance between continued health care innovation and public safety is imperative. We encourage NIH to ensure that scientists with expertise in human-technology interactions be involved in advisory bodies and other oversight arenas to help guide the efficacy and ethical practices of AI implementation and review.

¹ Matheny, M. E., Whicher, D., & Israni, S. T. (2020). Artificial intelligence in health care: a report from the National Academy of Medicine. *Jama*, 323(6), 509-510. <https://jamanetwork.com/journals/jama/article-abstract/2757958>; Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H. C., ... & Jeste, D. V. (2021). Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9), 856-864. <https://doi.org/10.1016/j.bpsc.2021.02.001>

² Fiske, A., Henningsen, P., & Buyx, A. (2019). Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *Journal of Medical Internet Research*, 21(5). <https://www.jmir.org/2019/5/e13216/>

Medical Ethics and Protecting Patients --APA agrees that new AI tools should centralize transparency and effectiveness in their development and deployment. Data being employed in the training of AI models raises serious privacy and transparency concerns. The effect and reach of AI tools is often unknown by the end user, fueling misunderstandings and misconceptions about the technologies; tech companies should take steps to increase transparency around the presence of these tools to ensure individuals are aware of the role AI might play in impacting their experience in real world and online environments. The use and exploitation of data will have unknown consequences on the perception and experience of AI tools.³ AI, like humans, are varied, yet people often see AI as homogeneous. This may result in AI failures being treated as global rather than specific, affecting perception and trust in AI.⁴ Trust between humans is earned. The human, the AI, and the context in which AI is employed determine trust in the technology, and the limits of this trust must be closely considered.⁵

Transparency requirements should be paired with public education efforts centered around technological literacy. NIH could take that work on or pass it to grantees, but it should not be neglected. Users should be informed about the way new AI technologies work and use user-provided data, how they might impact their lives, and how best to interact with these platforms to minimize negative impacts. Deployment of AI tools is often done without proper consideration of ethics and efficacy. AI must be properly tested prior to widespread deployment and the use of these technologies must be limited to applications where it is safe and effective, minimizing unknown, unintended, and inequitable impacts. These new AI technologies can be greatly improved by psychological science.⁶

User experience researchers and designers with strong backgrounds in psychological science can play an important role in ensuring new tools are more human, more effective, and more ethical. The data collection being fueled by the increased deployment of AI tools is leading to increased risks for individual data privacy. Tech companies must be required to redouble their efforts to protect data and ensure privacy around data collected by AI technologies, especially in the case of sensitive information possessed by health care providers and when copyrighted material is present. Companies must also be more transparent related to user data practices and data used for AI training. We hope this is an area in which NIH's size, scope and importance may help encourage private tech companies to be better partners for the extramural research community.

Goal 4: Support for a Federated Biomedical Research Data Infrastructure

While we realize that there are harmonization efforts occurring across federal agencies, we sometimes hear from our members that different agencies' requirements can vary. This can be a source of frustration, particularly for scientists who are analyzing data collected by multiple federal agencies.

Flexible or modifiable data representations can maximize interoperability and data integration across data sets, while still allowing for standardized definitions when needed. For instance, combining data

³ Sethumadhavan, A. (2019). Trust in Artificial Intelligence. *Ergonomics in Design*, 27(2), 34–34. <https://doi.org/10.1177/1064804618818592>

⁴ Longoni, C., Cian, L., & Kyung, E. J. (2023). Algorithmic Transference: People Overgeneralize Failures of AI in the Government. *Journal of Marketing Research*, 60(1), 170–188. <https://doi.org/10.1177/00222437221110139>

⁵ Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in Artificial Intelligence: Meta-Analytic Findings. *Human Factors*, 65(2), 337–359. <https://doi.org/10.1177/00187208211013988>

⁶ Gopnik, A. (2017). Making AI More Human. *Scientific American*, June 2017. https://www.scientificamerican.com/Papers_Alison/scientificamerican%20Gopnik0617-60.pdf

definitions of demographic categories in CMS vs SAMHSA data systems can be enhanced if the organizations' permit collection in ways that permit ease of alignment while still preserving individual organizational needs. Additionally, needs of individual, local, and state stakeholders should be considered at the same time federal needs are considered. Nuanced ability to aggregate and disaggregate demographic variables (such as race/ethnicity/sexual and gender minorities, disability, and rural/urban) will be critical to monitor potential health disparities among other topics. For example, one state may wish to disaggregate Native American vs Alaska Native data, and another state may wish to disaggregate Asian vs Pacific Islander, and data systems should be flexible to allow for both aggregate and disaggregated data. In similar fashion, there may be times where researchers, practitioners, or data consumers wish to align WHO/ICD/CPT coding with other systems such as DSM 5, and systems ought to be flexible enough to allow both integration and separation; this includes freely available coding manuals and data definition manuals, and cross walks such as to, for instance, align measures such as NCQA/HEDIS with SAMHSA URS or GPRA data as possible and appropriate.

Goal 5: Strengthen a Broad Community in Data Science

Developing this cadre of scientists is important. We agree that focusing on under-represented groups is the right emphasis. APA encourages the development of a training model that would emphasize data science training within programs like public health or behavioral medicine. Data science is ideally a cross-cutting discipline and adding some training in departments other than computer science is a good way to increase collaboration.

Again, APA extends its thanks for developing this thoughtful plan, and we welcome an opportunity to serve as a resource to your office in modifying and implementing it. Please let me know if the American Psychological Association may be of any assistance. If you have any questions or if APA can serve as a resource to your office, please contact Angela Sharpe, Sr. Director, Congressional & Federal Relations, at asharpe@apa.org.

Sincerely yours,



Mitchell Prinstein, PhD
Chief Science Officer
America Psychological Association