Association of
American Medical Colleges
655 K Street, N.W., Suite 100, Washington, D.C. 20001-2399
T 202 828 0400 F 202 828 1125
www.aamc.org

September 1, 2021

Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue NW
Washington, DC 20504

National Science Foundation
2415 Eisenhower Ave
Alexandria, VA 22314

**Re: Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource (NAIRR) 86 FR 39081**

*Submitted electronically to* https://www.federalregister.gov/documents/2021/07/23/2021-15660/request-for-information-rfi-on-an-implementation-plan-for-a-national-artificial-intelligence#open-comment

The Association of American Medical Colleges (AAMC) appreciates the opportunity to provide feedback to the White House Office of Science and Technology Policy (OSTP) and the National Science Foundation (NSF) on how to develop an implementation roadmap for a shared, national Artificial Intelligence (AI) research infrastructure, referred to as the 'National Artificial Intelligence Research Resource' (NAIRR). The AAMC is a nonprofit association dedicated to transforming health through medical education, health care, medical research, and community collaborations. Its members are all 155 accredited U.S. and 17 accredited Canadian medical schools; approximately 400 teaching hospitals and health systems, including Department of Veterans Affairs medical centers; and more than 70 academic societies. Through these institutions and organizations, the AAMC leads and serves America's medical schools and teaching hospitals and the millions of individuals employed across academic medicine, including more than 186,000 full-time faculty members, 94,000 medical students, 145,000 resident physicians, and 60,000 graduate students and postdoctoral researchers in the biomedical sciences.

It is our understanding that OSTP and NSF have issued this Request for Information (RFI) to inform the work of the NAIRR Task Force, which has been directed by Congress to develop an

implementation roadmap for a shared AI research infrastructure. We are pleased that this infrastructure will provide AI researchers and students across scientific disciplines with access to computational resources, high-quality data, educational tools, and user support.

The AAMC commends Congress' prioritization of AI, which has tremendous potential to advance human health and usher in a new era of biomedicine. In equal measure, the AAMC commends the aspirations of the OSTP and NSF to develop an AI infrastructure that serves as a resource for *all* citizens. As stated in the RFI, NAIRR is intended to "enable all of America's diverse AI researchers to fully participate in exploring innovative ideas for advancing AI, including communities, institutions, and regions that have been traditionally underserved." However, formidable barriers to the implementation of an inclusive AI research infrastructure are the long-standing and systemic discrimination, biases, and inequities that exist in the U.S. – all of which are present in the many overlapping sectors that converge upon the field of AI. Data demonstrate that the U.S. clinical and research enterprise is likewise marbled with biases and inequities, which, left uninhibited, preclude the formation of an *equitable* AI framework that, when operationalized, benefits all communities.

Nearly all aspects of AI – machine learning, deep learning, robotics, natural language processing, facial recognition, etc. – have huge implications for the nation's health. Radiology, telehealth, the "-omics" revolution (e.g., genomics), precision medicine, and personalized medicine are a few of the areas where the transformative impact of AI has already been felt. The AAMC leads and serves academic medicine to improve the health of people everywhere, and it is from this lens that we submit comments on the following section of this RFI:

*3. How the NAIRR and its components reinforce principles of ethical and responsible research and development of AI, such as those concerning issues of racial and gender equity, fairness, bias, civil rights, transparency, and accountability.*

Our comments underscore the necessity of building a NAIRR that 1) identifies and addresses systemic inequities at the interface of AI and biomedicine 2) mitigates bias by promoting representative datasets and algorithms 3) provides users with a data management and sharing plan that promotes community engagement and transparency, and 4) fosters a diverse AI workforce and leadership.

### ***Build a NAIRR That Identifies and Addresses the Impact of Systemic Inequities at the Interface of Biomedicine and AI***

Social determinants of health (SDOH)[1,2,3] – the conditions in which individuals are born, live, learn, work, play, worship, and age – have a profound impact on health. Copious amounts of research demonstrate that differences in SDOH account for significant and persistent health

[1] CDC National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP)
[2] Addressing Health Equity in Public Health Practice: Frameworks, Promising Strategies, and Measurement Considerations. Leandris C. Liburd, Jeffrey E. Hall, Jonetta J. Mpofu, Sheree Marshall Williams, Karen Bouye, and Ana Penman-Aguilar. Annual Review of Public Health 2020 41:1, 417-432
[3] AAMC Health Equity Research and Policy

inequities in the United States among groups that have and continue to be marginalized (e.g., racial and ethnic minorities, sexual and gender minorities, and individuals living in poverty). Such inequities are largely due to structural practices and policies that have systemically prevented communities from attaining their highest level of health and well-being. One prominent barrier is structural racism, which affects systems of housing, criminal justice, education, credit, and health care to reinforce discriminatory practices, beliefs, and distribution of resources. Of late, diverse stakeholders in the nation's biomedical enterprise – government, public health agencies, research and medical journals, non-profit biomedical societies and organizations, and academic institutions of higher education – have adopted policies to identify and address systemic inequities.

Currently, the playing field is vastly uneven, and examples of health inequities abound[4]. The ongoing COVID-19 pandemic, in which marginalized and disadvantaged populations continue to suffer greater morbidity and mortality[5], has highlighted the fact that scientific and medical breakthroughs and therapeutics do not equally benefit all individuals and communities. Left unaddressed, progress in biomedical advancements, which benefits from and is deeply connected to AI, is poised to selectively benefit a *subset* of communities – thereby amplifying and deepening health inequities. Foundational investments in AI – such as NAIRR – must therefore recognize how the differences in social advantages selectively affect which individuals and communities benefit from AI as well as engage in the AI workforce, education, and leadership. Interventions targeting SDOH and health inequities have tremendous potential to narrow the inequities in medicine, research, and AI. The AAMC urges the NAIRR to adopt an AI research framework that identifies, anticipates, and helps dismantle inequities. We strongly urge the NAIRR to partner with diverse communities in the development of this framework, its translation to infrastructure, and in subsequent and appropriate decisions regarding its use. Fundamentally, a framework built without community trust[6] cannot sustain meaningful and sustained diversity. While our examples have primarily focused on racial equity, we note the importance of intersectionality in properly addressing structural racism and discrimination. We hope that the OSTP and NSF will consider in its work all groups facing health inequities, including, but not limited to, racial and ethnic minorities, women, persons with disabilities, individuals in rural and underserved geographic areas, sexual and gender minorities, as well as other individuals from diverse backgrounds.

Our specific recommendations on creating such an infrastructure are as follows:

1. Establish a diverse council of community, patient, and scholarly experts to identify the mechanisms in which systemic inequities, biases, and discrimination (e.g., sexism, racism) in biomedicine impact AI research and development. Examples include the National Academy of Medicine's 'Digital Health Action Collaborative' and 'Culture Inclusion & Equity Action Collaborative', which respectively "foster improvements and innovation in digital infrastructure so that health technology is developed and applied in

[4] Agency for Healthcare Research and Quality: 2019 National Healthcare Quality and Disparities Report
[5] Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *JAMA.* 2020;323(24):2466–2467. doi:10.1001/jama.2020.8598
[6] AAMC Principles of Trustworthiness

ways that consistently lead to better population and patient-level health" and "advances a culture of health equity and engagement that places the needs of people and communities".

2. The council should 1) be comprised of individuals from diverse sectors and industries 2) be a standing component of the NAIRR infrastructure and 3) represent the broad spectrum of sciences, including biomedicine, clinical care, social sciences, and education.
3. The NAIRR should adopt an anti-racist agenda in concert with OSTP and NSF, as well as formulate practices that are explicitly anti-racist and oriented towards equity.
4. The NAIRR should fund and highlight research that examines how AI research and development is impacted by bias, discrimination, health inequities, and SDOH.

***Mitigate Bias in AI by Promoting Representative and Unbiased Datasets and Algorithms***

The power of AI lies in its ability to use what is known to predict the unknown[7]. AI systems rely on the data that they are given, which, as stated, often reflects various forms of bias. Data bias[8] is found in two primary forms – both of which are prevalent in biomedicine and can negatively impact aspirations of the NAIRR being equitable and inclusive. One form of data bias occurs when the data available is not an accurate representation of the reality that it is supposed to reflect. In this case, incomplete data gathering, faulty methodologies, and lack of reporting can lead to bias. This type of data bias has unfortunately been rampant in biomedicine. For example, it has been a common practice for biomedical researchers to use primarily male animals as research models, despite distinct, and often dramatic, sex-based differences in health and disease. Likewise, clinical trials – which are intended evaluate the impact of treatments on human health – often are characterized by a dearth of women and individuals from racial and ethnic underrepresented backgrounds. The National Institutes of Health (NIH) and our member institutions are actively addressing these matters.

Another prevalent form of bias occurs when the underlying processes of data collection promote inequalities. For example, machine learning, which is utilized by many investigators in biomedicine, uses algorithms to identify patterns through vast digital images, records, and data. Due to systemic and long-standing discrimination (e.g., racism, sexism, and homophobia), women and racial and ethnic minorities are underrepresented in medical data, resulting in algorithms that are based on data generated predominantly from white males. Though progress has been made in the inclusion of women and underrepresented minorities in datasets, a 2021 study[9] published in *JAMA* reported that from 2015 to 2019, the reporting of race and socioeconomic status (SES) in randomized clinical trials published in general medical journals was grossly underreported. Because AI uses vast amounts of "known data" to predict the unknown, lack of representation in data creates models that do not accurately reflect or accurately predict outcomes for a large portion of the U.S. population, thereby exacerbating discriminatory dynamics that promote social inequality. Lastly, we note that a lack of data altogether is another source that introduces bias.

[7] NAM: Channeling the Potential of AI to Advance Health Equity (Webinar Video)
[8] The AI Now Report: The Social and Economic Implications of Artificial Intelligence
[9] Alegria M, Sud S, Steinberg BE, Gai N, Siddiqui A. Reporting of Participant Race, Sex, and Socioeconomic Status in Randomized Clinical Trials in General Medical Journals, 2015 vs 2019. *JAMA Netw Open.* 2021;4(5):e2111516. doi:10.1001/jamanetworkopen.2021.11516

As revolutions in biomedicine, including big data, -omics, and deep learning, continue to synergize with AI, the AAMC urges OSTP and NSF to develop a framework that considers the impact of existing biases in AI systems. Below we offer four suggestions:

1. Addressing AI bias will require a systemic and multi-organizational approach. We recommend that the NAIRR standardize protocols that identify and mitigate forms of implicit and explicit bias in data and AI.[10,11]
2. Given the vast amounts of data, industries and applications that will converge with the NAIRR, we recommend that the NAIRR convene a panel to help codify and advise all relevant organizations and groups on strategies to address the many forms of data biases in AI. Such biases include[12], but are not limited to, algorithmic bias in translational research, data science, and machine learning; human bias; and data bias.
3. To foster reproducibility, transparency, and interpretation, the NAIRR should establish a policy whereby results, algorithms, and datasets show disaggregated data on intersectionality factors, such as race, gender, and SES.
4. We recommend that NAIRR fund research that develops interventions and address biases in AI, particularly at the intersection of AI and biomedicine.

***Foster Open Access of Data, Transparency, and Accountability***

The RFI states that a goal of the NAIRR is to create a "shared computing infrastructure to facilitate access to advanced computing resources for researchers across the country, including provision of curated data sets, compute resources, educational tools and services, a user-interface portal, secure access control, resident expertise, and scalability of such infrastructure." To realize such aspirations, the AAMC recommends the following:

1. The NAIRR should convene a council or body tasked with codifying requirements for accuracy, and fairness in ensuring the integrity of data, research, and protocols. We note especially that this council should not only capture 'scholarly experts', but also include patient and community members who bring diverse perspectives, expertise, and experience.
2. The NAIRR should adopt a policy of transparency and openness that fosters public trust and understanding in the uses, processes, and funding surrounding the NAIRR.
3. The NAIRR should establish a data access, management, and sharing policy, which outlines the shared expectation for transparency. This data management plan should also account for privacy, security, and legal matters, as well as the archiving of data. Amongst the research community, there is concern that AI manuscripts are irreproducible because authors commonly neglect to include details of their methods, parameters, and/or precise data processing steps. The AAMC recommends that all published work report their methods, process, and workflow, thereby enabling reproduction of original and subsequent analysis.

[10] 2021 National Academy of Medicine Webinar: Channeling the Potential of AI to Advance Health Equity
[11] Bias in Artificial Intelligence, Society to Improve Diagnosis in Medicine
[12] 2021 National Academy of Medicine Webinar: Channeling the Potential of AI to Advance Health Equity

4. The AAMC recommends requiring authors to publish disaggregated data on human subjects, including gender, and racial and ethnic background.

The AAMC has worked in partnership with NIH and other organization to promote the benefits of open sharing of data and the use of data repositories for managing federally funded research[13,14]. The AAMC believes that the open sharing of data in the NAIRR can promote the democratization and access of AI data by all communities.

***Create and Sustain a Diverse AI Workforce***

Historically, STEMM fields in academia have failed to advance women and underrepresented ethnic minorities[15,16] in faculty positions and positions of leadership. Moreover, data suggest that STEMM careers have seen uneven progress in gender, racial and ethnic diversity[17]. For example, various subfields of AI, such as computer science, are racially homogenous and dominated by men, curating a narrow segment of perspectives and experiences. If the NAIRR is to benefit all citizens, it must represent a *diverse* group of individuals – in leadership, data, education, and training. The AAMC therefore recommends that the NAIRR create a strategic plan[18] for broadening and incorporating the full range of perspectives, identities, and disciplinary backgrounds into all facets of the NAIRR.

In summary, the AAMC appreciates the opportunity to submit these comments that the NSF and OSTP have expressed as priorities. If you have any questions regarding this response, please feel free to contact me at rmckinney@aamc.org or Julia Omotade, PhD, Senior Specialist, Science Policy, at jomotade@aamc.org.

Sincerely,

Ross McKinney, Jr., MD
Chief Scientific Officer

---

[13] AAMC Comments on Data Repositories for Managing Federally Funded Research Data
[14] 2020 AAMC Response to the Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded Research
[15] National Science Foundation: National Center for Science and Engineering Statistics. (2021). *Women, Minorities, and Persons with Disabilities in* Science *and Engineering. https://ncses.nsf.gov/pubs/nsf21321/report*
[16] Lautenberger D, Dandar, V (2020). The State of Women in Academic Medicine 2018-2019: Exploring Pathways to Equity. Association of American Medical Colleges. Available at: https://store.aamc.org/downloadable/download/sample/sample_id/330/
[17] Fry R, Kennedy B, Funk C (2021). "STEM Jobs See Uneven Progress In Increasing Gender, Racial And Ethnic Diversity" PEW Research Center.
[18] National Academies of Sciences, Engineering, and Medicine. (2020). *Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors*. Washington, DC: The National Academies Press. Available at: https://doi.org/10.17226/25585.