Federal Register Notice 86 FR 46278, https://www.federalregister.gov/documents/2021/08/18/2021-17737/request-for-information-rfi-on-an-implementation-plan-for-a-national-artificial-intelligence, October 1, 2021.

Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource: Responses

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

Department of Ophthalmology
and Visual Sciences

September 29th, 2021

To:

The White House Office of Science and Technology Policy and National Science Foundation

RFI Response: National AI Research Resource

From:

Amitha Domalpally, MD, PhD, Department of Ophthalmology and Visual Sciences

Roomasa Channa, MD, Department of Ophthalmology and Visual Sciences

Pilar Ossorio PhD, JD, Professor of Law and Bioethics

University of Wisconsin, Madison

As researchers active in the field of AI in ophthalmology, we are appreciative of the opportunity to provide feedback to the National Artificial Intelligence Research Resource (NAIRR) Task Force. With grass root level experience in AI research, we are enthusiastic about the potential of AI to improve health systems and patient care. We are also aware of the obstacles in applying algorithms to patient care and can bring forth diverse perspectives to the implementation roadmap. We are hopeful that the guidelines provided by this task force will have direct impact and propel AI research for future researchers in many ways. Our specific comments are listed below:

2*. Which capabilities and services (see, for example, item D above) provided through the NAIRR should be prioritized?* *https://www.federalregister.gov/d/2021-15660/p-31*

The introduction of artificial intelligence (AI) to health sciences has transformed the detection and quantification of pathology from medical imaging bringing in a new era of diagnostics. The last 3 years have seen an explosion of research comparing clinicians' diagnosis to diagnosis by AI, with AI often showing superior diagnostic performance.

The initial excitement with AI has now settled into a phase of realistic expectations as limitations in performance of AI algorithms are recognized when they are implemented into real world clinical care. Editorial comments on algorithmic bias, lack of generalizability and real-world performance are trending. It took an enormous amount of research effort, time, and money for us as researchers to realize that AI does not have to be autonomous or unsupervised in order to favorably impact patients care and we are now ready for a pragmatic approach to AI. Our recommendations for prioritization include the following

1. **Resources for Study Design**: The pathway for FDA approval of therapeutics is well established with a phased approach involving defined goals, indications of use and a critical examination of results at every stage. Apart from basic concepts such as training and testing, AI related research is relatively unstructured. In many situations, with independent researchers, training and testing is performed using a limited dataset and if fortunate, ends in a cul-de-sac of publications. There is much AI research for the sake of hashtag AI. Education in the field of AI research in terms of study design, sample size, types of validation, data fairness and bias are needed to ensure that high quality research, that can be readily translated to improve health care, is conducted.

2. **Interdisciplinary Collaboration**: Collaboration between clinicians and AI developers is important at all levels of AI model development, particularly in identifying real world challenges and addressing them from the outset. Even companies such as Google that have access to unlimited data have had major setbacks when implementing their diabetic retinopathy algorithm. A comparison of multiple deep learning models for diabetic retinopathy uncovered disparities in real world performance.

3. **Availability of Standardized Imaging Datasets**: Ophthalmology has been a high yield research area for AI applications. Autonomous diagnostic system for diabetic retinopathy is one of the first FDA approved and one of the most promising AI tools in medicine. Ophthalmology is fortunate to have many publicly available datasets of ocular images for AI training. However, many of these have incomplete metadata and issues with image quality and annotations. Ocular image format is another area that severely lacks harmonization with proprietary formats rampant with all the manufacturers. Annotation is a difficult time-consuming task and crowd sourcing is not a practical option for medical imaging. Most research groups resort to the low-quality public datasets or develop their own training datasets, with unclear labeling methods. Testing is also performed on available retrospective data from clinical care with clinician diagnosis as reference standards. Even though this process is only a few years old, it still seems like the dark ages of AI research. **Standardized, well curated training**

and testing datasets are needed to prevent further disarray in the field of AI research.

4. **Creation of a Central Repository**: Many areas of AI research have central repositories for testing model accuracy, e.g., the Stanford Question Answering Dataset (SQuAd) or ImageNet. There is no such central repository in ophthalmology which results in multiple independent models being developed. Researchers work in silos and often do not share their algorithms. In addition, there are no pretrained algorithms that can be used. Providing a central repository offers opportunities for collaboration and prevents duplication of effort.

The NAIRR task force has rightly identified that a major overhaul is needed in the AI research process, particularly in the area of universal access to curated datasets. Creation of independent datasets of images and other clinical data that are considered benchmark, with constant improvement based on user feedback is imminently needed. While many of the general guidelines in medical imaging and AI can be implemented in ophthalmology, representation of ophthalmology as a field in the implementation roadmap is required to address the distinctive imaging requirements. Health Data Research of UK has taken this step and identified ophthalmology as a designated field https://www.insight.hdrhub.org/.

Ophthalmology is an image intensive field with many advanced imaging techniques that allow cellular level visualization. Ocular imaging equipment are specialized and developed specifically for imaging the eye. It is also a unique field of medicine where images are not interpreted by radiologists. Clinical images are interpreted by ophthalmologists and research images acquired as part of clinical trials are interpreted by certified graders (non-ophthalmologists) in reading centers. The latter is considered reference standard for FDA approval of AI algorithms. The eye has transparent tissue and allows direct visualization of blood vessels and neuronal tissue permitting retinal images to be used for predicting systemic diseases such as Alzheimer's and cardiovascular disease. AI research in ocular imaging is positioned to be a game changer and can provide a platform to provide a proof-of-concept sandbox that can be translated to the larger body of medical research.

3. *How can 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? https://www.federalregister.gov/d/2021-15660/p-32*

AI has the potential to dramatically improve diagnosis, prognosis, and image analysis. AI can leverage real-world data to complement and expand knowledge gleaned from clinical trials. However, to ensure that AI realizes its full potential in healthcare, the processes for developing and deploying it must be guided by and embody ethical norms. We define the ethical approach for AI research on three broad principles - non-maleficence, respect for persons, and justice.

Non-maleficence (do no harm): This principle is generally associated with ethical and regulatory requirements that researchers minimize risk of harm to data subjects. Research-related harms can arise from the processes of constructing and updating research datasets, or from the training and deployment of AI in healthcare. Historically, the research community has used anonymity of data subjects as the mechanism for lowering informational risks in research. Recently, there has been **extensive discussion about what counts as sufficiently anonymized or de-identified data and whether high dimensional health data (including whole genome or whole exome sequences) ever should be treated as de-identified.** The risk of re-identification increases as machine learning models become more sophisticated, computing power becomes more available, and datasets become more ubiquitous and higher dimensional. A dataset intended for widespread sharing should have utility for a variety of researchers and research projects over a fairly long time span. The variety of authorized uses of a dataset, and the unforeseen future uses, mean that both data subjects and researchers have a high degree of epistemic uncertainty concerning the potential harms or benefits that could arise from authorized uses, or from misuse.

Respect for Persons: This principle is prominent in discussions of **autonomy** and **informed consent**. Respecting persons requires that researchers promote participants' autonomy and informed consent is a means for doing so. Individual informed consent has been a mainstay of traditional clinical research. With the advent of AI, some scholars have argued that individual informed consent for research using electronic health records is impracticable (prohibitively time consuming and expensive), likely to bias results, likely to slow or hinder the development of AI tools, and unnecessary for low-risk research.

As both an ethical and regulatory matter, consent can be waived under some circumstances even though doing so may lessen the degree to which research promotes autonomy. In the U.S., researchers have never been required to seek consent or HIPAA authorization for using *de-identified* data collected in the clinical context. Despite the widespread and longstanding use of health record data for research, without individual informed consent, this practice is controversial as demonstrated by the ubiquitous negative response to the story of HeLa cells and by empirical studies. Black, Indigenous, and People of Color (BIPOC) have long been underrepresented in biomedical research, and this problem could be exacerbated because they have greater concerns about data sharing than non-BIPOC.

The 2017 revision of the Common Rule introduced a requirement for "broad consent" for the storage, maintenance, and secondary research use of *identifiable* private

information or identifiable biospecimens, including material collected in the clinical context. However, under HIPAA and the Common Rule whole-genome sequences and retina scans can be treated as de-identified even though they intrinsically individuate a person. One concern about requiring broad consent for the research use of clinical data and specimens is that health care institutions will need to track such consents for all their patients, and most institutions currently lack infrastructure and procedures for doing so.

Justice: Among other things, justice in AI research has to do with the fair distribution of the science's burdens, potential harms, and potential benefits; fair access to research, including fair processes for choosing research questions and for determining eligibility; minimizing pernicious bias to the extent possible, in both datasets and the algorithms trained on them; and maximizing the generalizability of algorithms to the extent reasonably possible.
Race, gender, and class biases have been well-documented in AI systems, including in AI for healthcare. Bias can occur at any stage of the AI development process and can lead to differential performance of algorithms for people from different social groups. BIPOC are often under-represented in data available for designing AI models, potentially resulting in models that are less effective or even dangerous for them. However, people can also be included inappropriately, represented in the data in ways that reflect social inequality existing outside of the healthcare system or pernicious bias within the healthcare system. In either of these cases, training and validating algorithms on such data can normalize, naturalize, or obscure unjust disparities.
In healthcare, when an algorithm produces results that differ by race, class, gender, sexual orientation, or other social category, it can be difficult to know whether this difference reflects the health consequences of living in an unequal society, or whether the difference reflects pernicious bias the algorithm learned from the data. If AI is designed with a "black box" approach, it is harder to identify or mitigate pernicious bias. Performance of AI should be evaluated in different settings and in different populations to evaluate for bias. Performance should be reported in terms of predefined metrics e.g., sensitivity and specificity stratified by population characteristics such as age, gender, race/ethnicity. It would be useful to have metrics for assessing the effects of pernicious bias in patterning health-related datasets and assessing how those metrics change in response to different strategies for acquiring and cleaning data.

Data Governance is key to responsible AI research. We define data governance as "a strategy for the overall management of the availability, usability, integrity, quality, and security of data in order to ensure that the potential of the data is maximized while regulatory compliance is achieved, and ethical norms are respected and integrated. Data governance helps a dataset comply with the FAIR principles. A governance strategy must be operationalized through organizational rules and norms, and structures such as steering and advisory committees. **Data governance can draw on a responsible innovation framework, such as AREA - Anticipation, Reflection, Engagement, and Action**. Under the AREA framework, people engaged in research

should consider the future consequences of their activities, develop and integrate explicit mechanisms for reflexivity, identify and engage relevant stakeholders, and respond to stakeholder feedback and self-examination by making changes to procedures and organizations (action). We believe the AREA approach would help ensure that ethical issues are contextualized and thoroughly analyzed rather than treated as compliance issues. Data governance using an AREA approach must extend across the lifecycle of a dataset, from data collection to processing, curation, sharing, use, and finally to the end of the life cycle (deletion). There may be ethical and technical issues unique to certain stages of that lifecycle, and other issues that recur across the lifecycle.

As stakeholders with direct involvement in AI research, we appreciate the opportunity to have a voice in this remarkable initiative. For any further clarification, please reach out to us.

Amitha Domalpally, MD, PhD,

Assistant Professor, Department of Ophthalmology and Visual Sciences

Research Director, Wisconsin Reading Center

Director, A-EYE Unit

University of Wisconsin, Madison

[redacted]

Dr.Roomasa Channa, MD

Assistant Professor, Department of Ophthalmology and Visual Sciences

Co-Director, A-EYE Unit

[redacted]

Pilar Ossorio PhD, JD,

Professor of Law and Bioethics

University of Wisconsin, Madison

[redacted]