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

Memorial Sloan Kettering
Cancer Center

8/31/2021

Attention:
Wendy Wigen, NCO
2415 Eisenhower Avenue
Alexandria, VA 22314

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

To whom this may concern:

As Director of Pathology Informatics and the current President of the Association for Pathology Informatics (API), I continually engage with digital and AI technologies along with their integration in healthcare. I share my personal opinions in response to the RFI on an implementation plan for the National AI Research Resource (NAIRR) in this letter.

Listed in the subsequent pages, I provide my responses for the Task Force to consider to all roadmap question items (A through I), underlined, and my reasoning. As part of my responses, I also give some opinions on how NAIRR and its components can reinforce principles of ethical and responsible research and the development of AI. Finally, I also include building blocks for the NAIRR, namely government, academic, professional organizations, and industry.

Again, I appreciate your attention.

Sincerely,

S. Joseph Sirintrapun, M.D.
Director of Pathology Informatics
Associate Attending
Department of Pathology
1275 York Avenue Rm A515, New York, NY 10065

Memorial Sloan Kettering Cancer Center
www.mskcc.org
NCI-designated Comprehensive Cancer Center

A. Goals for establishment and sustainment of a National Artificial Intelligence Research Resource and metrics for success;

My overarching goals are for 1) democratizing availability of healthcare data to AI community stakeholders, 2) sustaining a collaborative community environment that brings together all the engaged stakeholders for continual AI development, and 3) ensuring that clinical AI deployment is effective and ensures trust in its use

Gaps:

- No systematic study is underway for a sufficient level of "anonymization" to share data between institutions to aggregate for AI development
- No systematic research is underway to understand how to capture, organize automatically, and "index" healthcare data at scale, with much of the progress in capturing data through inefficient manual curation
- No practical/usable ontologic knowledge frameworks that engage both human domain experts and machines, and likewise, there are no systematic means to develop AI tools that work off these frameworks
- No strategic direction on the regulation of AI and how to handle "change control" or versioning of algorithms as they evolve
- No incentive to "operationalize" AI into clinical workflows with process engineering that will ensure trust in matching the AI intended to use to the clinical situation at the correct point in time.

Addressing gaps through funding and resource allocation directed towards:

- Creation of innovative and sustainable incentive models (i.e., ONC, HHS/CMS) for health institutions to anonymize, curate, and release healthcare data to cloud infrastructures
- Study the science of anonymization of different healthcare data types and how to make such data portable
- Study the art and science of anonymization for different kinds of healthcare data, and then build standardized tools that scale and are extensible
- Study the art and science of evaluating demographic and social determinant data, and then produce data models by which tools get built to capture, qualify, and quantitate biases and disparities
- Study the art and science of multimodal diagnostic data (i.e., clinical impression,

radiologic, pathology, and laboratory), and then build data models that enable the feature and label engineering for AI tools

- Study the art and science of patient status and clinical outcome multimodal data (i.e., clinical impression, radiologic, pathology, and laboratory), and then build data models that enable the feature and label engineering for AI tools
- Study "change control," or the handling in how updated versions of AI get deployed into clinical environments.
- Study of implementation science with AI, meaning how to deliver AI into clinical workflows; this means:
  - Deploying AI at the right time and situation with minimal friction
  - Optimized trust that the AI intended use case meets the clinical situation at hand in the correct context when triggered
  - Maintaining the right balance of "human in the loop" not to overburden yet ensure governance over AI tools to not becomes too autonomous without trusted oversight

B. A plan for ownership and administration of the National Artificial Intelligence Research Resource, including:

i. An appropriate agency or organization responsible for the implementation, deployment and administration of the Research Resource; and

The solution is not just one agency but a cross-agency partnership between HHS entities, including the FDA, HHS, and ONC. This partnership includes professional clinical organizations with AI expertise (i.e., The Association for Pathology Informatics, a member of the FDA Network of Experts). Also included in the partnerships are industry and big tech stakeholders, all in a collaborative environment to discuss the issues and solutions to overcome gaps and even provide resources.

ii. A governance structure for the Research Resource, including oversight and decision-making authorities;

A governance structure will require a cross-agency partnership between HHS entities, including FDA, HHS, and ONC, and in collaboration with professional clinical organizations with AI expertise (i.e., The Association for Pathology Informatics, which is also a member of the FDA Network of Experts). Also included in the partnerships are industry and big tech groups, in a collaborative environment to discuss the issues and solutions to overcome gaps and even provide resources.

C. A model for governance and oversight to establish strategic direction, make programmatic decisions, and manage the allocation of resources;

No such model exists and will need iteration as the cross-agency partnership with stakeholder engagement evolves.

D. Capabilities required to create and maintain 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;

May require engagement by the big tech cloud platforms (i.e. Amazon, Google) about computing infrastructures.

E. An assessment of, and recommended solutions to, barriers to the dissemination and use of high-quality government data sets as part of the National Artificial Intelligence Research Resource;

Bringing the multiple stakeholders together through coordination, leadership, and commitment with funding with issues that arise through the duration of the effort; all are requirements to ensure the sustainability of this effort.

F. An assessment of security requirements associated with the National Artificial Intelligence Research Resource and its management of access controls;

No comment, since I'm not a security expert. The science and "operationalization" of anonymization is critical in sharing data without exposure to sensitive information.

G. An assessment of privacy and civil rights and civil liberties requirements associated with the National Artificial Intelligence Research Resource and its research;

No comment, since I'm not a privacy/ethics expert. The science and "operationalization" of anonymization is critical in sharing data without exposure to sensitive information.

H. A plan for sustaining the National Artificial Intelligence Research Resource, including through Federal funding and partnerships with the private sector; and

May need big tech to participate through a "win-win" incentive partnership model for such a significant government/academic/vendor collaboration. This effort will require leadership with a solid strategic vision.

I. Parameters for the establishment and sustainment of the National Artificial Intelligence Research Resource, including agency roles and responsibilities.

No comment, only to state that such a resource should ensure that academic/professional organizations should get incentivized for sustained engagement in the collaborations.

1. What options should the Task Force consider for any of roadmap elements A through I above, and why? [Please take care to annotate your responses to this question by indicating the letter(s) of the item (A through I in the list above) for which you are identifying options.

See my responses above.

2. Which capabilities and services (see, for example, item D above) provided through the NAIRR should be prioritized?

See my responses for A above, where I list the gaps and my opinions on funding and resource allocation.

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?

- Study the science of anonymization of different healthcare data types and how to make such data portable
- Study the art and science of anonymization for different kinds of healthcare data, and then build standardized tools that scale and are extensible
- Study the art and science of evaluating demographic and social determinant data, and then produce data models by which tools get built to capture, qualify, and quantitate biases and disparities
- Study and develop better models in data use that ensures trust and ethical use for the generators of the data

4. What building blocks already exist for the NAIRR, in terms of government, academic, or private-sector activities, resources, and services?

One building block is leveraging partnerships with existing professional organizations like the Association for Pathology Informatics (API). API is the foremost organization with internationally recognized top experts in the field of pathology informatics. In addition, API has joined the FDA Network of Experts and the Network of Digital Health Experts. The FDA Network of Experts consists of a select group of 100 professional non-profit status organizations that identify relevant member experts and connect them to the FDA program members with timely feedback (not opinion or recommendation) on newly emergent health technologies. API, having experts in digital pathology and AI and synchronizing with the FDA Network of Experts mission, is a valuable existing resource for NAIRR.

5. What role should public-private partnerships play in the NAIRR? What exemplars could be used as a model?

Such partnerships are integral in advancing and incentivizing future sustained thriving development. One exemplar is the Defense Advanced Research Projects Agency (DARPA), seeking potential public-private partnerships to further the Robotic Servicing of Geosynchronous Satellites program.

6. Where do you see limitations in the ability of the NAIRR to democratize access to AI R&D? And how could these limitations be overcome?

See my responses for A above, where I list the gaps and my opinions on funding and resource allocation.