

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

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views and/or opinions of the U.S. Government nor those of the National AI Research Resource Task Force., and/or any other Federal agencies and/or government entities. We bear no responsibility for the accuracy, legality, or content of all external links included in this document.

09/30/2021

Subject: RFI Response: National AI Research Resource

We read with interest the RFI of the National Artificial Intelligence Research Resource (NAIRR) Task Force to develop a roadmap for a nationally shared research infrastructure. This initiative is expected to provide Artificial Intelligence (AI) and Machine Learning (ML) researchers and trainees across the scientific disciplines with access to computational resources, high-quality data, educational tools, and user support. We at Moffitt Cancer Center full-heartedly support such an initiative, as these resources are much needed to advance AI/ML in general, as well to tackle the chronic pandemic of cancer through proper application of this powerful technology.

On one hand, AI/ML has been presented as a savior technology that will transform oncology practice by automating laborious routine tasks, improving efficiencies, reducing costs, and enhancing decision-making support of complex oncology processes, including everything from treatment management to prediction of outcomes and adapting prescriptions over the course of treatment. On the other hand, the anticipated implementation of AI/ML in oncology and healthcare in general has been limited in scope and sometimes stagnant with less than 5% of major healthcare providers implementing any form of AI/ML solutions (HIMSS). Moreover, fairness and ethical concerns have casted shadows on the ability of AI/ML to reduce health disparities versus exacerbate existing inequities. Recognizing the potentials of AI/ML to transform cancer care and the need to address barriers in the realization of its potential, we have established the first AI/ML department in oncology at Moffitt Cancer Center, which was featured in the Cancer Letter ([Vol.47 No.07, February 19, 2021](#)). A key mandate of establishing this department is the integration with the parent cancer center scholarly activities in research and the facilitation of AI/ML clinical translation as part of the center's data/digital ecosystem, by enabling AI/ML driven multi-scale convergence of patient data into actionable cancer research and daily oncology care.

Given the aforementioned background, we provide our responses to the questions posed by the NAIRR's RFI, which we hope would be helpful in developing the proposed roadmap while highlighting specific cancer needs:

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.]*
 - A) The vision presented by the task group and the lessons learned for a secure single access platform and friendly presentation are welcomed. However, additional lessons learned from the TCIA/TCGA repository are that proper data structures and proper annotations for AI/ML are needed for successful utilization. In addition, indicators of data quality may be required when such data are used for training or testing of existing and/or new AI/ML algorithms and technologies.
 - B) Data governance and open access have been a challenge in the past, even through federally supported entities, including a multi-institutional cooperative groups, where democratization of data is [part of their mission, were unfortunately mired with regulatory

and technical challenges. Hence, it would be useful to have clearer and transparent criteria for resource sharing and data access eligibility and an independent oversight committee to organize such processes in a fair and transparent manner.

- C) The governance model needs to be representative of stakeholders of the different disciplines including sensitive sectors such as healthcare. We advise that an organizational structure should be created that can provide equal opportunity access given the expected high demand. This can be done in a similar fashion to how federal granting agencies currently provide resources. For instance, the application process can be structured in a hierarchal manner from the participating institutions through a nomination process to ensure that resources are properly allocated, and the use cases will have the expected impact on the targeted application area. Additionally, priorities can be provided within this process to meet national interests for AI/ML applications in the different sectors.
- D) The compute capability for developing and applying AI/ML already exists in the cloud-based platforms noted by the RFI. However, a major hinderance is the for-profit nature of many of these platforms; subsidization may be needed for academic and research use. As in response to A), curated data sharing is another major challenge; support for federated learning infrastructures may be necessary in the long run to enrich the available data resources while respecting legal and ethical concerns.
- E) Making high-quality, government funded datasets available is an important objective. However, there are several lingering challenges to this objective, which include the availability and open access due to privacy concerns and other regulation and technical issues. Supporting institutional infrastructures for secure federated data access may offer a feasible solution and facilitate resource sharing in an equitable manner.
- F) Proper cybersecurity safeguards would need to be evaluated and necessary requirements should be presented in a clear and transparent matter. This will provide the opportunity to replace antiquated security protocols with more modern approaches that can protect the data from unauthorized use without significantly hindering access and utilization.
- G) Privacy in healthcare is a major concern and standardization of protocols for anonymization and harmonization of the process would be an advantage for a centralized setting. In the federated setting, anonymization may not be necessary if proper security guards are put in place that can protect the data from unauthorized use while facilitating secure access.
- H) It is important to have sustainable infrastructure with long term objectives for hardware and software upgrades as the nature of AI/ML is to evolve over time and the need for better computing hardware and software are likely to persist and increase.
- I) The required parameters for sustainability need to meet the needs of the different sectors and their pressing use cases. This is especially true in the domains of healthcare and cancer research, where such needs continue to evolve over time. Moreover, a repository to maintaining codes, such as GitHub, with better security guards may be needed as well for such sharable resource too.

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

Provision of curated datasets has been a major limitation for widescale implementation and validation of existing and newly developed AI/ML tools at academic centers. ImageNet has been a driving force for the latest AI success, and something similar would be essential for benchmarking different algorithms across the different sectors especially in biomedical sciences. If too many resources are required for building an exhaustive AI/ML database, subsets of government funded and curated datasets could be sequestered for evaluations purposes only, which would still be very valuable for having common benchmarking schemes across the board for newly acclaimed AI/ML algorithms that can assess their robustness in a systematic fashion.

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?

There are tools, though still evolving, for evaluating AI/ML fairness that need to be encouraged and utilized as part of the AI/ML lifecycle. These tools include the interpretability and explainability evaluation tools. Moreover, the datasets provision of item D) can also be utilized as a testing bed of the fairness of the developed AI tools and how representative of the society and its underrepresented minorities. These are likely to be use case dependent but general principles can be chartered as suggested by the recent European Union and the FDA guidelines for trustworthy AI. In addition, specific datasets that include underrepresented populations may need to be generated as part of the implementation process for human biomedical research such that AI/ML algorithms trained/tested on those data are more broadly generalizable. Indeed, optimizing the generalizability of AI/ML to more diverse populations can accelerate the narrowing of the current disparity gap instead of exacerbating it as has been postulated

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

There are extensive compute resources, such as the ones presented by the STRIDE or the XSEDE supercomputer platforms, that can be leveraged and expanded to include more institutions. The main current limitation is in the availability of proper datasets for training and validation across the various domains, especially in the case of cancer research, which may limit the options for standardized vetting of newly proposed AI/ML techniques and subsequent translation into the clinic and daily care.

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

The STRIDE example, if it can be generalized, may be a good one to follow across multiple disciplines. However, affordable chargeback policies may need to be implemented with eth commercial vendors. Moreover, challenges associated with existing data commons such as the TCIA/TCGA repositories including proper data structures and quality metrics may need to be resolved for more effective utilization

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?

The limitations may not be technical but mostly related to proper access to datasets that would allow standardized evaluation of newly proposed AI/ML technologies along their lifecycle. Moreover, getting buy-in from institutions to make their existing resources available to NAIRR may be another challenge that may need to be handled through proper incentives and value-based propositions that can lead to the shared positive impact on the society and its institutions including cancer care.

On behalf of Moffitt Cancer Center, Tampa, Florida

Issam El Naqa, PhD, Chair of Machine Learning

Dana Rollison, PhD, Associate Cancer Director and Chief Data Officer

Edmondo Robinson, MD, Chief Digital Officer