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

Machine Learning Group
Physics Division
Lawrence Berkeley National Laboratory (LBNL)
1 Cyclotron Road
Berkeley, CA 94702

To whom it may concern,

The goal of the Machine Learning group in the Physics Division at Lawrence Berkeley National Laboratory is to advance the potential for discovery and interdisciplinary collaboration by approaching fundamental physics challenges through the lens of modern machine learning. Our group is a cross-cutting effort that connects researchers developing, adapting, and deploying artificial intelligence (AI) and machine learning (ML) solutions to fundamental physics challenges across the High Energy Physics frontiers, including theory. We are grateful for the opportunity to respond to the Request for Information (RFI) related to the National Artificial Intelligence Research Resource (NAIRR). We have discussed the questions posed to researchers and have some comments and suggestions related to the first question about the roadmap elements that the Task Force should consider.

For the administration of the NAIRR (B.a), we believe it is important to involve both domain scientists and computing/statistics experts. For educational tools and services (D), we think it is critical to support the development of educational resources at the undergraduate and graduate levels. Furthermore, we believe that there should be domain-specific resources. This is particularly important given the large burden of coursework and other training that many students have to take, in addition to the domain-specific requirements. A related topic is the support for common tool development. National Laboratories are particularly well-suited to host software developers for generic and scientific software necessary for AI/ML research and development as well as large-scale deployment (e.g. on High Performance Computing resources). These tools include open source software like scikit-learn, PyTorch, and TensorFlow. In addition to software support, national computing centers could also provide dedicated scientific on demand resources similar to Google's Colab.

The NAIRR specifically mentioned curated datasets (D) and we agree that this is critically important for innovation and development. It is essential that there be physical sciences-oriented datasets, which have unique challenges and thus also unique opportunities for method development that would not be possible with standard datasets like MNIST. In particular, the physical sciences make significant use of large scale *ab initio* simulations, which can become powerful tools for discovery when combined with AI/ML methods. There are also stringent requirements on uncertainty quantification. Additionally, many analysis methods from industry such as anomaly detection do not directly apply. Anomalies in many fundamental physics applications are "group anomalies" where no single example is anomalous and it is only on a statistical basis that one can declare a discovery. Many cutting edge experiments also have stringent requirements on AI/ML inference latency and environment robustness (e.g. to ionizing radiation).

While it is critical to develop and host curated, domain-specific datasets, it is also critical to ensure that there are strategies in place to assess the potential harms of AI/ML applications (G). In particular, there could be hidden costs or biases in AI/ML methods that require domain expertise and/or new AI/ML methodology to mitigate.

Thank you again for the opportunity and we would be happy to expand upon any of the above points.

Sincerely,

Dr. Benjamin Nachman ([redacted]
On behalf of the LBNL Physics Division Machine Learning Group
https://www.physics.lbl.gov/machinelearning