CRA Computing Research Association CCC Computing Community Consortium Catalyst AAAI Association for the Advancement of Artificial Intelligence

Computing Community Consortium (CCC), Computing Research Association-Industry (CRA-I), and The Association for the Advancement of Artificial Intelligence (AAAI) Responds to the Implementation Plan for a National Artificial Intelligence Research Resource

September 2021

Liz Bradley (CCC Chair and University of Colorado Boulder), Nadya Bliss (Arizona State University), William Gropp (University of Illinois Urbana-Champaign), Helen Nissenbaum (Cornell Tech), Chris Ramming (CRA-I and VMware), Ann Schwartz (CCC), Bart Selman (AAAI President and Cornell University), Helen Wright (CCC)

Response to Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource: https://www.federalregister.gov/documents/2021/07/23/2021-15660/request-for-information-rfi-on-an-implementation-plan-for-a-national-artificial-intelligence

The following is a joint response from the Computing Community Consortium (CCC), Computing Research Association-Industry (CRA-I), and the Association for the Advancement of Artificial Intelligence (AAAI). We offer this joint response to the following points from the RFI drawing from the extensive discussions within the national AI research community that arose while developing *A 20-Year Community Roadmap for Artificial Intelligence Research in the US* (AI Roadmap) as well as conversations among the newly formed CRA-I Steering Committee and community. All text which is bold and italicized is directly from the RFI. Text which is a regular font is the response.

1. ***What options should the Task Force consider for any of the roadmap elements A through I, and why?***
    a. ***Goals for establishment and sustainment of a National Artificial Intelligence Research Resource and metrics for success;***

        A NAIRR should support work on a range of core challenges in AI research and development. In our AI Roadmap, we identified three core themes central to the further development of AI:

        i. Integrated intelligence, including developing foundational principles for combining modular AI capabilities and skills, approaches for contextualizing general capabilities to suit specific uses, creation of open shared repositories of machine understandable world knowledge, and understanding human intelligence both to inspire novel AI approaches and to develop models of human cognition.

ii. Meaningful interaction, comprising techniques for productive collaboration in mixed teams of humans and machines, combining diverse communication modalities (verbal, visual, emotional) while respecting privacy, responsible and trustworthy behaviors that can be corrected directly by users, and fruitful online and real-world interaction among humans and AI systems.

iii. Self-aware learning, developing robust and trustworthy learning, quantifying uncertainty and durability, learning from small amounts of data and through instruction, incorporating prior knowledge into learning, developing causal and steerable models from numerical data and observations, and learning real-time behaviors for intentional sensing and acting.

Each of these areas requires a setting with interactive data collection, where machines and humans interact and machine learning is a dynamic process involving active learning in a changing environment. Static datasets can only provide a starting point for tackling these research challenges. Data collection and generation needs to be dynamic, guided by AI decision making itself. Examples are (1) the use of AI for scientific discovery, where AI methods guide the scientific experimentation in a continuous loop of data analysis followed by subsequent experimentation and new data collection and analysis, (2) work on human-robot interaction where the AI system uses active learning to explore and map its environment and dynamic interactions with users, (3) in AI for healthcare, where AI data analysis and decision support systems are part of the clinical process, continuously monitoring health data and providing continuous guidance, and (4) personalized AI for lifelong learning, where the AI education support system adjusts to new learning goals and needs over time.

The AI Roadmap therefore recommends a broader perspective on the required shared National AI Research Infrastructure. Such infrastructure would embed AI research in actual application environments. One example would be an "AI-ready research hospital" where AI researchers from around the country could work (likely remotely) with clinicians and other medical staff to develop AI for health applications. The AI systems would provide data analysis and input to the clinical process, while continuously monitoring patient progress and multiple sources of health data. Another example would be a shared materials science research facility that enables the development of AI-driven automated experimentation searching for new materials. A final example would be a shared facility for the development of interactive robots for assisted living. Active learning and continuous data collection are again central to the development of the next generation of assistive robots.

These examples all point to the need for shared AI Research Infrastructure resources that can provide active continuous data collection in dynamic environments to train the next generation of AI decision support systems. The AI systems need to be integrated into these environments to allow for active learning (where the AI system itself decides what data to collect next), and for dynamic control and decision making.

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… ?***

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

In looking at the organization and management of the NAIRR, it is important to recognize both the key importance of data and the reality that that data will be distributed across the country, in commercial clouds, at national research facilities, and at academic institutions. An NAIRR needs to place adequate computing near the data, which implies that there are multiple sites providing computing resources for NAIRR. This is not a problem, and in fact offers some advantages. Unlike Leadership Computing Facilities, it isn’t necessary to have all of the computing in one place, and few single jobs would need to use all of the computing resources at once in a tightly-coupled way – and research that needs such resources could instead use DOE Exascale systems through the DOE INCITE program. An advantage is that with multiple sites, and with both public and private providers, the NAIRR can include a broad menu of different computing technologies and ensure that there are frequent updates and incentives for exploring innovative new hardware and software.

NSF has a successful model for this approach. NSF funds advanced computing systems, including innovative pilots, and provides a virtual organization through the Extreme Science and Engineering Discovery Environment (XSEDE). The flexibility of this organization was recently demonstrated in the COVID-19 HPC Consortium, where XSEDE played a key role in managing the review of proposals and allocations of the resources, which included contributions from commercial clouds, national laboratories, and academic supercomputer centers. While XSEDE is ending next year after many successful years of operation, NSF is continuing this virtual organization model through a new program, ACCESS. While the details of the service model used in XSEDE and ACCESS will be different for NAIRR, the use of a virtual organization following the XSEDE model provides a tested, successful model for providing access to research computing while ensuring ongoing innovation, and supporting a distributed model essential for providing access to the many disparate data sources. One difference from the XSEDE model is that NAIRR should also have regular, planned investment in both hardware and software, something that is not part of XSEDE (which only provides support services – hardware and software are funded through separate competitive NSF solicitations).

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;***

   i. Create something similar to the National AI Research Centers, which are intended to create unique and stable environments for large multidisciplinary and multi university teams devoted to long-term AI research and its integration with education. A shared computing infrastructure should be:

      1. Funded through decade-long commitments, to provide stability and continuity of research

2. Multi-stakeholder centers with a smaller core set of partners and a large network of affiliated educational institutions, national labs, and industry. Promoting collaboration between academia, industry, and government will enable cross-cutting research and technology transition.
3. Multidisciplinary in the expertise involved in the research areas
4. Multi-faceted in their research goals
5. Effective dissemination vehicles for significant results

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;***

i. Access to data sets is a huge problem, and this issue is not limited to government data. The field has matured beyond its initial academic focus on algorithms and theories and into a context of continuous data collection, social and interactive experimentation, and massive amounts of knowledge about a constantly changing world. Building from those foundations, the tech industry has compiled and leveraged massive resources—datasets, knowledge graphs, special-purpose computers, and large cadres of AI engineers—to propel powerful innovations. It is important to create *open* AI platforms and resources, which will be a vast interlinked distributed collection of "AI-ready" resources (such as curated high-quality datasets, software, knowledge repositories, testbeds for personal assistants and robotics environments) contributed by and available to the academic research community, as well as to industry and government.

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

i. Shared resources (as highlighted in other sections of this response, such as XSEDE and DOE resources) have implemented security requirements and those requirements should be consulted for NAIRR. NAIRR shared resources should as a baseline follow all recommended shared resources requirements of other research infrastructures.

ii. There are additional considerations in context of AI research around data, results, systems, adversarial AI techniques, and domain specific security. The following described what is meant by each:

1. Data: well curated data sets from a vast variety of sources are vital to AI research. At the moment, academic researchers do not have access to the same quality of data as industry researchers. A significant benefit of implementation of this resource would be providing access to datasets that the researchers may not have access otherwise. However, different datasets would likely require different restrictions and those will need to be managed through this implementation plan (administratively, legally, technically, etc). Secure multi-party computation, homomorphic encryption, and hardware trusted execution environments (TEEs) can potentially overcome barriers to confidential data sharing and should be considered as long and short-term foundations for the creation and use of of relevant datasets
2. Results: given the potential sensitivity of AI algorithms and system results on particular datasets, it will be necessary to make sure that the requirements that are applied to datasets are also applied to the derivative products of those datasets.
3. Systems: as per (1) and (2), a key benefit of the NAIRR would be the ability to seamlessly combine various datasets and results into novel AI systems. The interactions between various components of those systems could potentially introduce either cybersecurity vulnerabilities or dataset biases. Thus, it is important that those are considered as part of access controls and software review. This is another area where potentially NAIRR could learn from other shared resources and/or “app store” implementations (for example, researchers may want to “publish” their algorithms and software modules for other researchers to use, but that “publication” process should include some assessment of potential security vulnerabilities).
4. Adversarial AI: an important area of research is understanding vulnerabilities in AI algorithms. It would be beneficial to the research community if NAIRR could provide a “sandbox” environment for adversarial AI research with appropriate ethical and security restrictions and potential assessment of national security implications.

5. Domain-specific security: given that AI research is performed and can be beneficial to a broad range of application domains, NAIRR should consider if additional security requirements need to be implemented in certain situations. For example, security and access requirements would likely be different between analysis of human genome data and development of AI algorithms for autonomous vehicles. As such, domain experts should be brought in for those situations.

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

   i. For the NAIRR, it is important to remember that all data is sensitive, especially in this AI and Machine Learning environment, all data can implicate all other data.
   ii. Data is a fundamental resource. Due to the enormous interest in all aspects of human and social life, data about people is highly valued. This inevitably raises privacy issues throughout the data cycles of collection, assembly, and analysis.These issues need to be grappled with not as an afterthought, for example, but at the moment of deployment or use, but as an integral consideration at each phase.
   iii. Access to all data needs to be broadened. Whenever you have tasks that are “handed off” to AI systems from a human actor, there also needs to be accountability, justice, and fairness.
   iv. Finally, if we are going to be investing ethical and political rights/values in a National AI Research Resource, we need to include research into values/rights so they apply to novel environments. For example, what does privacy mean? What does transparency mean? What does autonomy mean?

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

   i. We need to partner with industry and academia to train highly skilled AI engineers and technicians. Extensive adult education programs, distance education programs, and online courses should be developed to fit personal circumstances and schedules of those interested in pursuing AI careers. Strong programs that attract students to AI from early stages (high school and undergraduate) and promote AI careers will be a key ingredient for growing a workforce with advanced AI expertise and sustaining the NAIRR.

2. ***Which capabilities and services (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;) provided through the NAIRR should be prioritized?***

    A. Open AI platforms and resources: a vast interlinked distributed collection of "AI-ready" resources (such as curated high quality datasets, software, knowledge repositories, testbeds for personal assistants and robotics environments) contributed by and available to the academic research community, industry, and government.
        a. Provide the AI community with substantial experimental resources for both basic research and applications.
        b. Reduce redundancy of effort and cost in the research enterprise, so research projects do not have to build up capabilities or collect new data from scratch each time.
        c. Reduce the cost of individual research programs to integrate relevant capabilities or to compare their work with others.
        d. Reduce cost of large teams of collaborators by providing already integrated or easy-to-integrate infrastructure.
        e. Provide the AI community with open resources that will bolster not only research in all of academia, but also in small technology companies, companies in other sectors, and government organizations.
    B. Incentivize emerging interdisciplinary AI areas: initiatives to encourage the research community to work in interdisciplinary AI studies—e.g., AI safety engineering, as well as analysis of the impact of AI on society—will ensure a workforce and a research ecosystem that understands the full context for AI solutions.
    C. Address AI and the future of work: these challenges are at the intersection of AI with other disciplines such as economics, public policy, and education. It is important to teach students how to think through the ethical and social implications of their work.
    D. Train highly skilled AI engineers and technicians: support and build upon the National AI Infrastructure to grow the AI pipeline through community

colleges, workforce retraining programs, certificate programs, and online degrees

   a. Develop AI curricula at all levels: guidelines should be developed for curricula that encourage early and ongoing interest in and understanding of AI, beginning in K-12 and extending through graduate courses and professional programs.
   b. Create recruitment and retention programs for advanced AI degrees: including grants for talented students to obtain advanced graduate degrees, retention programs for doctoral-level researchers, and additional resources to support and enfranchise AI teaching faculty.
   c. Engage underrepresented and underprivileged groups: programs to bring the best talent into the AI research effort.

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?

A. It is not just research and development that needs ethical attention, but the deployment and use are equally, if not more, important. It is imperative to incorporate ethics and related responsibility principles as central elements in the design and operation of AI systems.
   a. Make sure that AI systems align with human values and norms to ensure that they behave ethically, taking into account potential risks, benefits, harms, and costs. In order to do this, AI systems will have to incorporate complex ethical and commonsense reasoning capabilities that are needed to reliably and flexibly exhibit ethical behavior in a wide variety of interaction and decision making situations.

B. An AI system must be able to explain its rationale to the team members (e.g., why it suggested certain experiments) and it must make its level of uncertainty clear in a way that team members can truly understand. Given the critical applications and outcomes of modern AI , these systems must also act reliably.

C. Explicitly codifying best practices in ethical behavior, conduct, and inclusiveness in academic, industry, and government organizations, so that inappropriate interactions, isolation, and implicit biases are eliminated from the school and the workplace.

D. Novel computing technologies often improve our lives, but they can also affect them in ways that are harmful or unjust. It is important to teach students and practitioners how to think through the ethical and social implications of their work.

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

A. One building block that already exists for the NAIRR is the National Science Foundation (NSF) CloudBank, a cloud access entity that helps the computer science community access and use commercial clouds for research and education by delivering a set of managed services designed to simplify access to commercial clouds

B. It is important to recognize that industry and academic work to support multicloud and hybrid cloud scenarios, since they will be enduring for some time (e.g. because of data sovereignty and privacy issues, to prevent lock-in to single-hyperscaler solutions, to take advantage of specialized cloud attributes, and to recognize the need to exploit local computing for cost and latency optimization).

C. As discussed in a recent CRA-Industry roundtable it is important to recognize that commercial clouds provide capabilities that could not be rivalled by government-funded alternatives. To avoid wasteful investment, alternative strategies need to be considered.

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

A. The NSF CloudBank suggests one direction for public private partnerships, which is mainly to get the hyperscalers to actively contribute or subsidize lower-cost services for government and academic research. And per above, there is another kind of public-private partnership that would help drive initiatives in multi-cloud/hybrid cloud access that would help prevent single-cloud lock-in and help perpetuate dual-source options for government/academic procurement.

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?

A. The US education system currently makes use of computing technologies, some of which are enhanced by AI, but there is still great room for improvement of these technologies, and even greater opportunity for full adoption to enhance our education system if access were free for all on all technical levels. If students had free access to high quality education, lifelong education and training would only enhance AI research and development. We could train the next generation of AI specialists, data scientists, and software engineers.

B. Another limitation, in relation to clouds, is that commercial clouds are highly democratizing in the sense that organizations that previously didn't have the resources to build their own computing facilities could now leapfrog those wealthier organizations by immediately and directly leveraging state-of-the-art hyperscale resources.

In conclusion, all of these responses came from conversations among the CRA committees and the AI Roadmap, which was the result of a community activity to articulate AI research priorities for the next 20 years in a wide range of areas in AI and related disciplines. The NAIRR, like the AI Roadmap, has the potential to bring the field to a new era of audacious AI research to tackle long-standing and multidisciplinary problems. These investments will significantly accelerate the development and deployment of AI technologies with a profound impact across all sectors of society.