MEETING SUMMARY
National Artificial Intelligence Research Resource Task Force
Meeting #5

February 16, 2022

MEETING SUMMARY

Meeting Summary

The fifth meeting of the National Artificial Intelligence Research Resource (NAIRR) Task Force (TF) was held online via Zoom on February 16, 2022, 11:00 AM–6:00 PM EST.

Welcome and Administrative Remarks

The meeting started at 11:05 AM EST.

Dr. Lynne Parker, NAIRR TF Co-Chair, opened the meeting. Dr. Manish Parashar, NAIRR TF Co-Chair, motioned to approve the summary from the prior NAIRR TF meeting; the motion passed. Dr. Parashar then introduced the agenda.

The meeting had five primary goals:

1. Consider options for security controls; privacy, civil rights, and civil liberties requirements; and technical integration of resources;
2. Consider sustainment and opportunities for public and private partnerships;
3. Hear from prospective users of a NAIRR on designing the NAIRR to meet user needs;
4. Discuss an approach to evaluation, metrics, and indicators of success for the NAIRR; and
5. Finalize an outline for the interim report.

In addition, time was reserved at the end of the meeting to address questions from the public, with any unanswered questions to be advanced to the beginning of the next NAIRR TF meeting. Dr. Parashar noted that no unanswered questions had been carried over from the previous meeting that was held on December 13, 2021.

The session ended at 11:16 AM EST.

Readout and Discussion of Draft Recommendations: User Access Controls and Usable Security

The session started at 11:16 AM EST.

Elham Tabassi, NAIRR TF member, presented the findings and proposed recommendations of the User Access Controls and Usable Security Working Group (WG), beginning with a recap of some insights extracted from responses to the TF’s public Request for Information (RFI). The WG found that NAIRR system elements will have heterogeneous security needs and that existing security processes and policies developed through extensive prior public and private efforts can be leveraged for the NAIRR. Ms. Tabassi noted that the NAIRR will need to keep up with new developments in cybersecurity, a rapidly-changing field, and that security risks are as much human as technical. The WG assumed that the NAIRR would provide access through a single sign-on, accommodate both open science and projects that require restricted access (e.g., those working with sensitive data), and that compute and data resources for a given project would reside on the same platform. The group recommended a tiered access model that leverages FedRAMP standards to define security requirements—without requiring FedRAMP certifications or approvals. They also recommended that the NAIRR have a dedicated, expert technical security staff, and provide regular and continuous hands-on training and security-related support accessible at all hours to all users and staff. The WG concluded by highlighting a few remaining open questions for further consideration.

TF members discussed the presentation. WG members described the “Five Safes” framework for data protection and the need to adjust from a culture of downloading data to having the data reside where the compute is to prevent exfiltration of sensitive data. TF members noted that not all aspects of the NAIRR will require high levels of security. WG members agreed, commenting that their approach was to plan in advance for the greatest security needs, with the expectation that the NAIRR would also make non-sensitive, unrestricted data available to users. TF members suggested updating the language to be used in the report to clarify that FedRAMP approval will not be required for all aspects of the NAIRR. The TF also discussed strategies that support strong user awareness of and compliance with security policies. Other discussion topics included limits on the level of sensitivity of data used within the NAIRR environment and the importance of balancing security and usability.

The session ended at 12:10 PM EST.

Readout and Discussion of Draft Recommendations: Privacy, Civil Rights, and Civil Liberties

The session started at 12:10 PM EST.

Dr. Parashar presented the recommendations of the Privacy, Civil Rights, and Civil Liberties WG, noting that the WG consulted with additional outside experts and leveraged input received via the public RFI and the panel discussion at the December 13, 2021, TF meeting. The WG found that the NAIRR has an opportunity to build strong governance frameworks, including standards for data governance and stewardship that could be widely adopted. Assessment of AI systems and tools must consider the broader social, political, and historical contexts in which they are developed and deployed. Dr. Parashar outlined a proposed NAIRR strategy for protecting privacy, civil rights, and civil liberties rooted in transparency and oversight. In this model, the NAIRR would establish a framework for vetting the appropriateness of research conducted via the NAIRR and associated outcomes; create mechanisms to support security and privacy requirements, enable monitoring, and ensure compliance; require regular reporting; and train users and resource providers about policies, responsibilities, and best practices. An ethics review board would have purview over relevant policies and assessments.

TF members discussed the presentation. One member suggested that the recommendation on training could be more nuanced. TF members discussed the potential to carve out a less-rigorous requirement to support use cases involving short-term, non-sensitive use of the NAIRR, such as a short classroom exercise to model a gravitational physics problem; utilizing existing research requirements, training, and vetting mechanisms; and tapping into the data science ethics curricula being developed in light of the growth of data science within academia. In addition, TF members discussed vetting of datasets, noting that assessment and validation of datasets would be an ongoing process for NAIRR. For example, another WG member commented that the NAIRR could vet datasets for biases and potential harm, such that inclusion in the NAIRR becomes a proxy stamp of approval. It was also suggested that the NAIRR could catalogue problematic datasets to be deprecated, and curate "biased-by-design" datasets to advance research on AI bias and to test models for robustness. TF members also discussed options for removing racist, bigoted, and otherwise toxic or offensive datasets from the NAIRR, noting the importance of vetting data labels. Dr. Parashar noted that notions of privacy and civil rights will evolve over time, highlighting the importance of maintaining transparency about NAIRR datasets, policies, practices, and decisions.

The session ended at 1:05 PM EST.

Break: 1:10–1:40 PM EST

Readout and Discussion of Draft Recommendations: Technical Integration

The session started at 1:40 PM EST.

Dr. Mike Norman, NAIRR TF member, presented the recommendations of the Technical Integration WG, beginning with the design assumptions used as a basis for WG deliberations and noting that the . WG also sought information from individuals with expertise in unified access portal development and edge computing. Overall, the WG determined that technical integration of federated compute resources is mature, while integration of AI including machine learning (ML) data repositories, edge computing resources, and AI testbeds is less mature and will require additional attention as the ecosystem develops. The WG recommended the NAIRR embrace existing standards, de facto standards, and best-of-breed open-source solutions to nurture an AI ecosystem while avoiding "one-off" integrations. To provide users with the latest tools and capabilities, the NAIRR resource pool should be refreshed frequently based on user needs, trends, and technological advances. To allow for a seamless and intuitive user experience across a spectrum of users, the WG recommended that the NAIRR user portal be designed with "walk-up tooling" for scientists with consistent user experiences across private, multi-, and hybrid cloud infrastructures. The portal should also support alternate access methods (e.g., shell, scripting) for more advanced users. To effectively integrate data repositories and edge computing devices, the WG recommended the NAIRR establish a network of exemplar ML data repositories with powerful search and retrieval capabilities, in addition to encouraging the development of standard edge computing middleware. Usage and allocation of NAIRR resources should be measured in U.S. dollars or dollar equivalents. Finally, the WG recommended that the NAIRR should be staffed with an AI/ML DevOps team in addition to the staff who would be needed at a high-performance computing center (HPC). In addition to staff who serve as resource providers as mentioned, the NAIRR should staff a user support center with AI training specialists, portal developers, and data analysts.

TF members discussed the presentation. One member suggested that the pricing model could be a challenge. Other members discussed how pricing has been handled in other contexts and suggested that a number of organizations understand how to monetize and amortize the costs of such an operation relatively well. A member noted that the pricing model may be included in the interim report. Members also discussed how the allocation system would work for disadvantaged groups, and if any unique steps would be taken with respect to organizational overhead fees including indirect costs. Several options for provisioning and administering services through the NAIRR were discussed, including organizational overhead, pass-through, and service-center models. A TF member brought up the issue of interconnects and the expense of installing them to effectively use resources across the NAIRR. The TF members discussed the current state of this issue, with cloud providers actively adopting HPC interconnects, and converged architectures being deployed. These interconnects should be part of the resource design, TF members stressed.

Members of this WG also answered questions from the public posed in the Q&A portal before the session ended, including budgeting for the NAIRR's capacity across the broad user base, pricing models, the use of container-based architectures to manage workflows, the implementation of privacy-enhancing technologies within the NAIRR, and potential data-sharing agreements.

The session ended at 2:42 PM EST.

Briefing: Public-Private Partnerships and Sustainment Considerations for the NAIRR

The session started at 2:42 PM EST.

Dr. Lisa Van Pay, Dr. Emily Grumbling, and Ms. Morgan Livingston (Science and Technology Policy Institute) presented findings from their review of options for public and private partnerships to help achieve the vision of the NAIRR, along with advantages and limitations of different types of partnerships for resource provision and meeting stakeholder needs. For compute, the core NAIRR management entity could build (or contract) new hardware or partner with federally-funded, high-performance, distributed, and/or other computing resources sponsored by agencies such as the National Science Foundation (NSF) and U.S. Department of Energy, or those available from commercial cloud computing service providers. Similar options exist for data infrastructure, along with leveraging open-source tools. Partnering with established entities would enable a speedier launch of the NAIRR system, would likely be less costly in the near term, and could help to provide a variety of options and experiences to users. Commercial resources have the advantage of more closely tracking the cutting edge of technology, though leveraging them could reinforce reliance on the private sector for AI R&D, risk vendor lock-in, and require negotiation of clear intellectual property and user policies. Leveraging open-source data tools could provide cost savings and support for the open-source ecosystem, but require active vulnerability and quality management.

Partnerships with different data holders could bring resources to NAIRR users that a NAIRR management entity would likely be unable to generate or sustain on its own. Federal agencies hold reliable data on pressing societal needs but have strong protection and compliance requirements. Private-sector entities likely hold large volumes of data well-suited to AI R&D, although they may not have strong incentives to share them. Universities and non-profits could offer research data but may lack sharing mechanisms or policies. In lieu of creating new AI educational resources, the NAIRR could partner with higher education or non-traditional education and training programs to link or provide access to their tools through the

NAIRR user portal, in service of broadening the AI talent pool. Beyond resource provision, partnerships can be leveraged for engagement with stakeholder communities, including civil-society groups and individual experts, to inform the NAIRR design and support its oversight—for example, in areas such as equity, privacy, civil rights, and civil liberties protections. Successful partnerships build on shared interests among entities and require alignment of incentives and clear legal agreements. Appropriate partnership mechanisms will depend on the types of entities involved.

The session ended at 3:00 PM EST.

Panel: User Perspectives on the NAIRR

The session started at 3:00 PM EST.

Panel speakers were:

- Tom Dietterich, Distinguished Professor Emeritus in the Collaborative Robotics and Intelligent Systems Institute, Oregon State University;
- Susanta Ghosh, Assistant Professor in Mechanical Engineering-Engineering Mechanics, Michigan Technological University;
- Kinnis Gosha, Hortinius I. Chenault Endowed Associate Professor of Computer Science, Morehouse College;
- Gail Rosen, Professor of Electrical and Computer Engineering, Drexel University;
- Rima Seiilova-Olson, Co-Founder and Chief Machine Learning Scientist, Kintsugi; and
- Carlos Theran-Suarez, Instructor, Computer and Information Sciences Department, Florida A&M University.

Dr. Parker introduced the panel, and each panelist spoke for five minutes. Panelists discussed approaches to allocate compute resources, the compute needs of startups, and training resources. Multiple panelists noted the need for higher amounts of startup allocations for compute than are currently available: researchers need sandbox time to familiarize themselves with new architectures, estimate the compute needs for their research, and establish initial results. A panelist also discussed how startups need free or near-free compute during their first two to three years of existence for the intense development needed to establish minimum viable products and attract initial investment. In addition to the need for more startup allocations than, for example, those provided on current NSF-funded resources accessible via the eXtreme Science and Engineering Discovery Environment (XSEDE), there is also a need for a portfolio approach to fund more experimental research. For example, the NAIRR may provide a unique opportunity to support the reproducibility of AI research, which is highly needed but often not funded by federal funding agencies such as NSF.

A panelist also cautioned that reliance on existing federal funding mechanisms for NAIRR allocations could increase the disparities in access to science funding: researchers without allocations will be less competitive for grants than those who have been awarded access and can use the NAIRR to conduct research leading to stronger future proposals. Another panelist commented on the need for inclusiveness:

the NAIRR should include Minority-Serving Institutions, single-gender and liberal arts schools, users spanning different education levels, and a diversity of department types. A panelist commented that it is important for the public to understand the NAIRR's impact and its benefit to society.

Discussion also touched on training needs: while online resources—such as Coursera and FastAI courses, GitHub repositories, and Hugging Face—are helpful for AI training, there is a need for an interactive component. Groups traditionally underrepresented and non-technical groups especially need community to code together, teach one another, and inspire one another. The NAIRR could incorporate a community-building component, such as through Slack or Discord.

The session ended at 4:01 PM EST.

Discussion: Public-Private Partnerships and Sustainment Considerations for the NAIRR

The session started at 4:01 PM EST.

Dr. Parker moderated a discussion among TF members on public and private partnerships and sustainment considerations for the NAIRR. A TF member suggested that partnerships would likely be needed to provide the variety of resource elements desired for the NAIRR in the absence of very high funding levels. Such partnerships would require potentially complex legal agreements, for example, to define intellectual property or data rights. In response to a question from the public, the TF considered the possibility that partners might be able to provide some resources as in-kind contributions, and noted there might be material benefits for doing so beyond direct financial compensation—for example, access to talent and data in exchange for providing compute and data storage.

One TF member suggested the TF might not need to decide what entities are included as resource providers as selection may be made via a formal request for proposals, and that economic considerations are important for managing supply and demand for the NAIRR. Another member suggested that the NAIRR will need a clearly defined business model to establish appropriate legal frameworks and governance, and to avoid conflicts of interest. Dr. Parker noted the TF would continue its discussions on this topic as it moves from the interim report to the final report.

The session ended at 4:14 PM EST.

Break: 4:14–4:25 PM EST

Discussion: Defining Indicators of Success for the NAIRR

The session started at 4:25 PM EST.

Dr. Parashar presented on indicators of success for the NAIRR. He identified desired outcomes of the NAIRR in terms of four high-level goals: innovation, diversity, capacity, and ethics. He proposed specific recommendations, including that the NAIRR management entity capture data to enable regular assessment of performance indicators and tracking of progress toward intended outcomes, and to inform responsive management decisions. Sufficient funds for these activities should be budgeted, and the NAIRR system could be designed to capture desired metrics readily. In particular, data gathering about the AI

R&D community should begin as soon as possible to provide an effective baseline for measuring change, drawing from data available through the National Center for Science and Engineering Statistics and the Computing Research Association's annual Taulbee survey. Dr. Parashar proposed that these indicators include measures of investment, resource usage, outputs, and impact, to be complemented by surveys to gauge user needs and satisfaction and capture researcher-level outcomes. Researchers should be provided with standard language to acknowledge the use of the NAIRR in their publications to enable tracking of outputs. Finally, Dr. Parashar presented a list of example data and metrics to gather.

TF members discussed the presentation and options for metrics. One member concurred with the four pillars proposed and noted that the NAIRR should be equally successful in all of them. In response to a question about evaluation, the group discussed the importance of an independent evaluator for achieving results and the need to assess performance relative to a counterfactual (i.e., the next highest value alternative to the chosen investment). A TF member commented that it was not clear whether the metrics proposed fully mapped to the pillars named, and the group discussed other sources of data, such as information gathered by universities or partner resource providers. Another TF member noted that performance evaluation is necessarily limited to the data available, and relying on quantitative metrics can result in decision making that targets improvement in metrics rather than the outcomes for which those metrics are a proxy. This challenge could be mitigated by changing which metrics are emphasized over time and by leveraging qualitative metrics. Regardless of how progress toward outcomes is measured, the path by which these are expected to be achieved (i.e., a theory of change) must be articulated.

The session ended at 5:01 PM EST.

Discussion: Interim Report Outline and Next Steps in Recommendation Finalization

The session started at 5:01 PM EST.

Dr. Parker presented a proposed outline for the interim report and tentative structure for the content to be included in each section. The report outline largely maps to the elements prescribed in the legislation that established the TF. TF members followed up with clarifying questions about where specific content elements fit into the outline and to confirm there would be an executive summary. One member commented that many options were presented by Science and Technology Policy Institute (STPI) researchers over the course of the TF's efforts to date, and wondered when decisions would be made about these. Dr. Parker commented that these options were generally presented in advance of the launch of the relevant working groups, and were considered as WG recommendations were formulated.

The group discussed how the interim report would focus on the "what"—clarifying the defining characteristics for the NAIRR—and the final report would elaborate on the "how," that is, a roadmap with concrete implementation steps. A TF member asked whether there would be an opportunity for stakeholders to provide feedback on the interim report; Dr. Parker noted that a second RFI would be released as the interim report is published to solicit public input on the draft. TF co-chairs will also reach out for feedback from technical experts at federal agencies.

Dr. Parker then described the timeline for completing the draft interim report, noting that STPI will be converting the WG outputs into a discussion draft, and that the document would be completed iteratively. Dr. Parker requested that TF members comply with quick-turn requests for feedback on each draft to ensure the TF meets its statutory obligation of delivering the interim report in May.

The session ended at 5:30 PM EST.

Questions from Public and Meeting Close

The session started at 5:30 PM EST.

Dr. Parashar moderated the TF in addressing the remaining question, about the potential for counterfactuals to be biased, submitted by public attendees via Zoom's Q&A portal.

Dr. Parashar concluded the session, thanking members of the TF, NSF, OSTP, STPI, and the public, and reminding everyone that meeting summaries, slide presentations, and details about upcoming meetings can be found at https://www.ai.gov/nairrtf/.

The next meeting is scheduled for April 8, 2022, at 11:00AM–5:00PM EST. Details will be posted to the Federal Register shortly.

The meeting adjourned at 5:35 PM EST.

Appendix I: Attendance for NAIRR TF Meeting #5

TF Members Present:

Manish Parashar, National Science Foundation (Co-Chair)

Lynne Parker, White House Office of Science and Technology Policy (Co-Chair)

Daniela Braga, DefinedCrowd

Mark Dean, retired (formerly IBM and University of Tennessee, Knoxville)

Oren Etzioni, Allen Institute for AI

Julia Lane, New York University

Fei-Fei Li, Stanford University

Andrew Moore, Google

Michael Norman, University of California, San Diego

Dan Stanzione, University of Texas, Austin

Frederick Streitz, Department of Energy

Elham Tabassi, NIST

TF Members Absent:

None