

Federal Register Notice 87 FR 31914, <https://www.federalregister.gov/documents/2022/05/25/2022-11223/request-for-information-rfi-on-implementing-initial-findings-and-recommendations-of-the-national>, May 25, 2022

Request for Information (RFI) on Implementing the Initial Findings and Recommendations of the National Artificial Intelligence Research Resource Task Force: Response

Anthropic

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views 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 the responses and external links included in this document.

ANTHROPIC

June 29, 2022

Submitted electronically via NAIRR-responses@nitrd.gov

Reference: 87 FR 31914, Document Number 2022-11223

Subject: Anthropic Comment regarding “Request for Information (RFI) on Implementing Initial Findings and Recommendations of the National Artificial Intelligence Research Resource Task Force”

Anthropic welcomes the opportunity to provide feedback to the Office of Science and Technology Policy (OSTP) and the National Science Foundation (NSF) in response to a Request for Information (RFI) on the initial findings and recommendations contained in the Interim Report of the National Artificial Intelligence Research Resource (NAIRR) Task Force.

Anthropic is an AI safety and research company working to build reliable, interpretable, and steerable artificial intelligence (AI) systems. We’re an organization with backgrounds in research, engineering, and policy, and we approach AI development from a cross-disciplinary perspective. Since our founding in early 2021, we’ve raised \$700 million, primarily to fund the buildout of world class infrastructure for running large-scale AI training jobs. For Anthropic, infrastructure is fundamental to the success of our organization and we invest significant amounts of capital and headcount into developing and maintaining it.

As articulated in our previous submission¹, we believe progress in AI safety and technical advancements depends on broad public participation in AI research. Unfortunately over the past several years, frontier AI research and development (R&D) has become heavily concentrated within a small number of corporate entities. The NAIRR represents an opportunity to democratize the resources required to build advanced AI systems, allowing non-commercial actors to actively participate in the R&D ecosystem and build systems that more accurately reflect the goals of society at-large.

We appreciate the opportunity to share our perspective on the Interim Report, specifically the proposed scope of the NAIRR, its resource requirements, and organizational design. At a high level, we recommend the NAIRR Task Force consider the following suggestions as it refines its plans for the Final Report:

¹ Anthropic. (2021, October). *Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource: Responses*. ai.gov. <https://www.ai.gov/rfi/2021/86-FR-39081/Anthropic-NAIRR-RFI-2021.pdf>

- Putting forward a more ambitious, detailed technical proposal for the scope of the NAIRR, in order to seize this opportunity to deeply invest in the United States’ AI research capacity **(Topic A & Topic B)**
- Distributing the majority of compute resources to a broad range of academic researchers, meanwhile allocating a non-trivial portion for research endeavors that rival industry-scale development efforts **(Topic B & Topic C)**
- Using the NAIRR not only as a repository of existing testbeds, but also as an opportunity to further increase the capacity to measure, assess, benchmark, and monitor AI systems **(Topic C)**
- Triaging NAIRR capabilities so that implementation can happen in a staged way – start with the minimum viable product (“MVP”) version of a NAIRR and iterate from there, while continuously testing for usability and efficiency of NAIRR infrastructure **(Topic B)**

A Bolder Vision for the NAIRR (Topic A) & A Concrete Technical Proposal (Topic C)

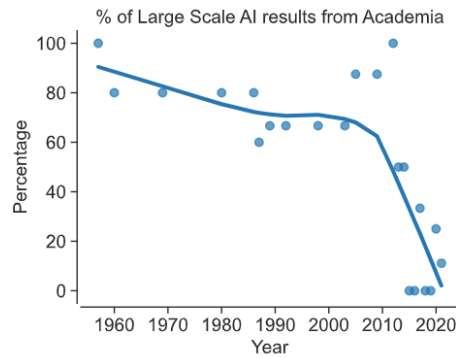
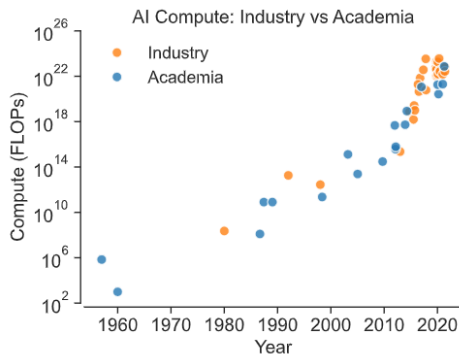
We echo the Task Force’s findings that potential breakthroughs in sustainability, national security, and other societal challenges depend on expanded access to AI resources, including compute hardware and sufficient data. The NAIRR represents an enormous opportunity for the U.S. Government to continue its longstanding support of academic research and advance AI development in the United States. While the Interim Report accurately notes the growing divide between academia and the private sector in resource accessibility (and as a result, opportunity to develop advanced systems), the compute divide is even more pronounced than what is alluded to in the report.

Compute-intensive research – the kind required to build more general-purpose, natural language and computer code processing models – is incredibly costly, and far exceeds the standard academic budget. For example, the training cost of OpenAI’s GPT-3 was estimated to be several million dollars², while Google’s PaLM was estimated to cost between \$9 million and \$23 million dollars³. As these models continue to grow in size and capabilities, the compute required to train them also increases and most academic institutions can’t allocate funding for resources at this scale. As a result, we’ve seen the industry contributions to large-scale AI research dwarf those of the academic community over the past decade⁴.

² Lambda. (2020, June 3). *OpenAI’s GPT-3 Language Model: A Technical Overview*. LambdaLabs. <https://lambdalabs.com/blog/demystifying-gpt-3/>

³ Heim, Lennart. (2022, April 5). *blog.heim.xyz*. Estimating PaLM’s training cost. <https://blog.heim.xyz/palm-training-cost/>

⁴ Ganguli, D., et al. (2022). Predictability and Surprise in Large Generative Models. arXiv. <https://arxiv.org/abs/2202.07785>



(Left) The amount of compute required by major AI projects over time is increasing exponentially for both academic (blue) and industrial (orange) projects. **(Right)** The proportion of large-scale AI results from academia is steadily decreasing. Blue curve represents a Lowess fit to the data⁴.

To meet the ambitious goal of the NAIRR to “strengthen and democratize the U.S. AI innovation ecosystem,” the U.S. Government must meet the moment with an equally ambitious and detailed infrastructure proposal. **Recommendation 3-1 of the Interim Report suggests multiple Federal agencies fund the NAIRR cooperatively, but does not point to specific agencies or detailed funding requirements.** Leaving this fundamental aspect of the NAIRR undefined leaves its potential success up to chance. We understand that funding recommendations may not be in scope for the Initial Report, but we believe it is a critical aspect to plan for in advance. We have included some estimated figures for consideration, and we urge the Task Force to include concrete recommendations and implementation plans in the Final Report. Ideally, these recommendations would be supplemented with draft legislation, which could be included as an appendix to the Final Report.

To assist in this effort, the Task Force may look to similar models implemented in Australia (National Computational Infrastructure⁵) and Canada (Advanced Research Computing (ARC) Platform⁶) to benchmark its own recommendations against the technical investments made in peer countries. However, the scope of these investments should be treated as a baseline, as year-over-year researcher demand continues to exceed available supply. In the case of Canada’s ARC Platform, only 24% of the total GPUs (a fundamental hardware component to today’s large-scale AI systems) requested were awarded to researchers in 2022 due to resource limitations⁷. If the United States wishes to sustain its leadership in AI research, it must invest in ways to make AI research more accessible. Outlining an ambitious and detailed technical proposal in the Task Force’s Final Report can be a step towards that effort.

⁵ NCI Australia. HPC Systems. <https://nci.org.au/our-systems/hpc-systems>

⁶ Digital Research Alliance of Canada. (2022). *Advanced Research Computing*. <https://alliancecan.ca/en/services/advanced-research-computing>

⁷ Digital Research Alliance of Canada. (2022). *2022 Resource Allocations Competition Results*. <https://alliancecan.ca/en/services/advanced-research-computing/research-portal/resource-allocation-competitions/2022-resource-allocations-competition-results>

Our vision of a truly competitive NAIRR for the United States would be on the order of a 100,000 GPU cluster, with an estimated cost of roughly \$4 billion for three years. This would require a significant financial investment and would represent an exemplary resource built over the course of several years. In our previous comment we contextualized this cost with the \$97 billion spent on capital expenditures⁸ by the leading U.S. digital infrastructure providers (Amazon, Google, Microsoft) in 2020. Private sector investment continues to grow: in 2021, those companies spent over \$124 billion on capital expenditures, representing an increase of 28% over the previous year⁹, and further widening the divide between resources available in industry and academia.

To align resource needs with an eventual appropriations recommendation, the Task Force may consider polling academic research departments across the United States to better understand current resource constraints and future research ambitions. Alternatively, the Task Force could partner with a think tank or non-profit organization focused on the United States' R&D ecosystem to carry out the polling exercise. In either case, we recommend the Task Force include a concrete technical proposal, informed by the needs of researchers, in the Final Report. Doing so will not only reveal the AI research potential of the academic community, but inform concrete steps towards building a stronger AI innovation ecosystem.

Create Opportunities for Large-Scale AI Experimentation (Topic B & Topic C)

Without public intervention along the lines of the NAIRR, access to compute resources will continue to be a research barrier between academia and industry. We agree with Recommendation 3-11 that resource allocation processes should be as inclusive as feasible, and **recommend the Task Force consider an additional funding and access tier for a small number of compute-intensive research projects.** As currently envisioned, the NAIRR will enable new AI research from a broader set of stakeholders, but the compute gap between academia and industry will continue to widen without a substantial investment in a handful of large-scale projects.

While the vast majority of resources would be made widely accessible to a broad range of researchers, something on the order of 30% of the NAIRR's computational capacity could be awarded to research efforts seeking to build industry-scale AI systems. Given the fact that compute-intensive models are developed almost exclusively within industry, and further investigation and access to those models is tightly controlled by corporate actors, **providing an opportunity for academic researchers to build and investigate equivalent systems directly**

⁸ Note that this CapEx figure includes spend on land, corporate offices, warehouses, etc., in addition to cloud and data center infrastructure.

⁹ Fitzgerald, C. (2022, February 16). *Follow the CAPEX: Cloud Table Stakes 2021 Retrospective*. Platformonomics. <https://www.platformonomics.com/2022/02/follow-the-capex-cloud-table-stakes-2021-retrospective/>

supports Recommendation 3-13 (“...resource allocation should be designed to incentivize contributions to the NAIRR user community or to the public good”).

We support the Task Force’s recommendation that all users of NAIRR computational resources pass a research proposal evaluation process (Recommendation 3-10), and **recommend that compute-intensive projects undergo an enhanced review that explicitly considers the potential societal impacts of such research**. To implement Recommendation 3-13 on a practical level, the proposal review process might also take into account the researchers’ intent to contribute such systems back to the NAIRR ecosystem, thereby enabling further investigation and experimentation by a broader community.

The NAIRR as a Catalyst for More System Testing (Topic C)

We view testbeds as a critical and underinvested facet of AI R&D, and echo the Task Force’s findings that robust testing can spur innovation, assess the safety of AI systems, and draw on the expertise of a broader range of academic disciplines. Of the Task Force’s definition of testbeds, Anthropic has primarily focused on the use of “data sets and frameworks that support evaluation,” though we view the NAIRR as a much-needed resource to also host testing environments. In addition to the benefits outlined in Findings 4-10 – 4-13, a rigorous testing framework can assist model developers, the broader research community, and the general public in evaluating AI systems for performance and safety. Specifically, testbeds can help with:

- Internal Benchmarking: When building models, developers can use testbeds to understand how particular implementation decisions affect model performance and safety.
- External Benchmarking: Developers, the broader research community, and the general public can use testbeds to compare the relative performance and safety features of models from different organizations side-by-side.
- Resource Forecasting: Developers can use testbeds to more predictably anticipate future resource needs and investment decisions.
- Understanding Societal Impacts: Developers, the broader research community, and the general public can use testbeds to evaluate models for important societal impacts (e.g. fairness, bias, and alignment with human values).

Finding 4-11 accurately notes that testbeds can increase equitable involvement in AI research by including less well-funded institutions – this can be expanded to include the opportunity that testbeds create for bringing in new perspectives and academic disciplines. We believe the most promising AI research will come from cross-disciplinary collaboration, not only in model development, but in model testing and validation, as well. **Unlike model development, investigating existing large-scale models is a relatively low-cost way to involve the diverse and varied expertise of the academic community.** For example, the Centre for the Governance

of AI estimated that a research project examining bias in GPT-3 required less than \$100 in compute resources¹⁰. At that cost, the NAIRR could potentially support thousands of researchers across the U.S. in carrying out impactful and socially-relevant work.

Recommendation 4-19 of the Interim Report (cataloging AI testbeds in the NAIRR) will help the broader research community identify and access the methods of testing that, as of today, are spread across the research literature, development community, and open source code repositories. Supporting this effort with a dedicated, full-time staff (as proposed in Recommendation 4-17) is an excellent use of NAIRR funds and will help keep these resources up-to-date and accessible. A catalog of testbeds will provide an excellent foundation for a more expansive vision of the NAIRR in measuring and assessing AI systems. **By classifying the kinds of testbeds within the catalog, the NAIRR staff will be able to identify gaps in existing testing methods and highlight areas for additional investment.** These findings can then be used to encourage researchers to develop *new* testbeds in particular domains, which the NAIRR can fund through its research proposal review process.

Implement the NAIRR in a Staged Approach (Topic B)

We are encouraged by the ambitious goals stated by the Task Force and the potential for the NAIRR to become the leading shared infrastructure for AI R&D. To ensure its success, we recommend the NAIRR development staff triage desired capabilities and services, and implement them in a staged way. While the NAIRR will ultimately provide for a range of infrastructure, data, and support options to a varied group of stakeholders, we recommend starting with the minimum viable product (“MVP”) version of a NAIRR that can help accelerate research at the outset.

As we suggested in our previous comment to the Task Force, **the first phase of the NAIRR should prioritize resources that are critical to research and readily-available**¹¹. These include: access to cloud infrastructure, a repository of easily accessible datasets, a catalog of existing testbeds, and sufficient funding for engineering support staff. While this first iteration could support a wide variety of research efforts, **it is essential that one of the initial goals of the NAIRR be to establish infrastructure that can support the provisioning of at least one large experiment**, where “large” involves more than 100 accelerator chips (e.g. GPUs) running in parallel to train a single machine learning (ML) model¹². Because infrastructure behaves

¹⁰ Anderljung, M., Heim, L., & Shevlane, T. (2022, April 11). *Compute Funds and Pre-trained Models*. Centre for the Governance of AI. <https://www.governance.ai/post/compute-funds-and-pre-trained-models>

¹¹ Anthropic. (2021, October). *Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource: Responses*. ai.gov. <https://www.ai.gov/rfi/2021/86-FR-39081/Anthropic-NAIRR-RFI-2021.pdf>

¹² To put this in perspective, a handful of recent research projects from industry developers used several times the number of GPUs proposed here. Meta recently released a model trained on over 900 A100 GPUs (“OPT: Open Pre-trained Transformer Language Models” - <https://arxiv.org/abs/2205.01068>), while a research effort from OpenAI

differently at large scales relative to small scales, funding dedicated DevOps and systems administration staff at the outset will help get researchers up and running with large-scale projects. Running such workloads early in the life of the NAIRR can serve to quickly validate the effectiveness of the infrastructure and identify areas for further improvement.

Alongside measures of infrastructure efficacy, the NAIRR staff should continuously evaluate other performance and usability metrics to inform subsequent phases of development. Staff may consider regular user polls or automated metrics that capture whether the NAIRR expands access to AI R&D resources (e.g. geographic representation of grant recipients), whether the NAIRR meets the resource needs of researchers (e.g. percentage of requests that are met), and its overall usability (e.g. number of days to launch a project after funding). With these insights, the NAIRR staff can then work to add in complementary resources such as hybrid infrastructure solutions, previously unreleased government datasets, and more complex testing environments.

Conclusion

We applaud the work of the Office of Science and Technology Policy, National Science Foundation, and the Task Force to develop a shared research ecosystem and encourage more equitable participation in AI R&D. Anthropic firmly supports the goals of the NAIRR and sees it as a tremendous opportunity to support critical and underfunded research into this transformative technology. Advancements made possible by the NAIRR could add to the long legacy of foundational research in academia that eventually powers technological innovation across the U.S. economy.

We urge the Task Force to put forward an ambitious vision for the scope of the NAIRR and funding, drawing inspiration from other “Big Science” infrastructure investments in fields such as physics and astronomy. We also recommend the Final Report include explicit proposals for compute allocations that can support industry-scale model development, as well as funding and development opportunities for new AI testbeds. We appreciate the opportunity to share our feedback and are eager to continue supporting the Task Force as it prepares its Final Report.

used over 700 V100 GPUs (“Video PreTraining (VPT): Learning to Act by Watching Unlabeled Online Videos”- <https://cdn.openai.com/vpt/Paper.pdf>).