CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

"The Main Resource is the Human"

A SURVEY OF AI RESEARCHERS ON THE IMPORTANCE OF COMPUTE

DATA BRIEF

AUTHORS
Micah Musser
Rebecca Gelles
Ronnie Kinoshita
Catherine Aiken
Andrew Lohn

APRIL 2023

Executive Summary

Artificial intelligence is increasingly understood as a strategic technology that governments seek to promote domestically and constrain for adversaries. One approach to promoting, or constraining, AI progress centers on the role of computational power (or "compute"). This approach encourages policymakers to enact policies and provide support to make compute resources more accessible to domestic researchers, perhaps while limiting their availability for strategic competitors. Relatedly, there are concerns that the increasing computational demands of AI breakthroughs risk concentrating AI research in the hands of a small number of well-resourced actors, limiting the diversity of AI researchers, what research receives meaningful attention, and who benefits from AI progress.

Despite growing policy attention to these issues, whether and how AI researchers view compute as a critical resource for their research is unknown. To address this gap, the Center for Security and Emerging Technology (CSET) surveyed more than four hundred AI researchers to examine their compute use, how they think about compute's role in AI progress, and the degree to which they are constrained (or not) by compute. Key findings include:

1. **Surveyed AI researchers are not primarily or exclusively constrained by compute access.** More respondents report talent as an important factor for project success, a higher priority with more funding, and a more limiting factor when deciding what projects to pursue. Data availability is also cited as a more common reason for rejecting projects.

2. **There are few differences between academic and industry AI researchers in terms of compute use and concerns.** While academic researchers report spending less money than industry AI researchers on compute, they report similar levels of hardware use. Both groups report similar levels of concern about insufficient compute allowing them to make meaningful contributions to AI research in the future.

3. **Academics report that changes in their compute needs outpace changes in their access more often than industry researchers.** However, most academics do not cite compute resources as a major factor that could cause them to leave for industry jobs.

4. **High compute users are more concerned about compute access.** Researchers reporting higher levels of compute use also report higher levels of concern about a lack of compute allowing them to make contributions, and more often select additional compute as a top budget priority.

5. **Surveyed AI researchers hold a range of opinions about government-provided AI research resources.** Most researchers select grant funding as a resource that would be useful to them, though many also select compute. Some researchers express skepticism about the government provision of these resources and concerns about an exclusive focus on scaling up compute.

Our results suggest that compute cannot be viewed as an all-purpose lever for promoting AI progress. Provisioning compute domestically and restricting access to it internationally may promote or constrain certain types of AI research, but may have less impact on other AI research areas. Well-intended attempts to democratize AI research by provisioning large-scale compute may even run the risk of exacerbating existing inequalities in compute use. This report's findings suggest that in some ways, talent is more important than compute for fostering AI research, so policymakers should evaluate how compute-focused interventions can be coupled with policies to foster AI talent in order to effectively promote AI research progress.

Table of Contents

Introduction

There is growing concern around possible inequities in artificial intelligence researchers' access to computing resources (or "compute") and resulting effects on the development of AI within the United States. This concern is summarized in the National Security Commission on Artificial Intelligence's *Final Report* in 2021, which concluded that—due to the compute and data needs of cutting-edge AI systems—"the development of AI in the United States is concentrated in fewer organizations in fewer geographic regions pursuing fewer research pathways."[1] Citing several findings regarding the rapid growth in compute demands for the largest deep learning models, the NSCAI argued that an increasing divide between the "haves" and "have nots" incentivizes academics to leave for industry jobs, undermines the competitiveness of startups, reduces diversity among AI researchers, and restricts the range of promising research avenues considered by scholars.[2]

Meanwhile in October 2022, the Biden administration announced new export controls on the sale to China of high-end GPUs.[3] The stated justification behind these controls was the need to prevent China from using advanced computing technologies "to produce advanced military systems including weapons of mass destruction . . . and commit human rights abuses."[4] Commentators, however, were quick to speculate that the export controls were also motivated by a desire to maintain a competitive advantage over China in fundamental AI progress.[5] If correct, this reflects a belief on the part of policymakers that computing power represents the most effective (or convenient) lever by which the United States can constrain AI progress in rival nations.

The growing emphasis on these two ideas—that compute is of central importance to AI progress and that researcher access to compute is increasingly stratified—is evident in recent proposals for a National Artificial Intelligence Research Resource.[6] In January 2023, the NAIRR Task Force submitted its final report to Congress and the president. The report calls on Congress to allocate $2.6 billion in funding for the NAIRR over the next six years, with $2.25 billion paid out in contracts to resource providers, where "the largest awards should be reserved for large computing investments."[7] The proposal views compute as central to both "spur[ring] innovation" and "increas[ing] diversity of talent" in AI research, and similar arguments are found in a variety of strategy documents and analytical reports.[8]

However, the degree to which AI researchers actually feel constrained by their access to compute is understudied. Discussions of the importance of compute and access to it are often framed around the compute demands of large or "cutting-edge" deep learning models, but do these discussions reflect the concerns of the broader

community of AI researchers?[9] Recent research finds that since 2012, publications from elite universities in top AI conferences and journals have crowded out researchers from less prestigious universities, which may be in part due to stratification in compute access.[10] But to what degree do AI researchers feel their work is constrained by a lack of access to compute?

This report addresses these questions by surveying AI researchers about their compute usage, level of concern regarding their future compute access, and the extent to which compute—as compared with other factors such as data availability or talent—limits the projects they work on. We find that compute is not the primary constraint faced by

We find that compute is not the primary constraint faced by many AI researchers, but that access to data or talent more directly constrains research plans and researcher behavior.

many AI researchers, but that access to data or talent more directly constrains research plans and researcher behavior. We also find little evidence that industry researchers use more compute than researchers in academia, or that academic researchers are more concerned about their level of access to compute. Respondents express a mix of views regarding the concept of the NAIRR, with general support for national AI research resources, but concerns about implementation.

The results presented here do not necessarily suggest that recent policy actions such as the proposed formation of the NAIRR or the imposition of export controls on high-end GPUs are misplaced, as access to compute is a bottleneck for some researchers.* In light of these results, however, this report suggests that policymakers temper their expectations regarding the impact that restrictive policies may have on computing resources, and that policymakers instead direct their efforts at other bottlenecks such as developing, attracting, and retaining talent.

---

* In particular, researchers working on "foundation models" or other highly compute-intensive research projects may be most constrained by compute. Compute-focused policymaking may disproportionately influence these subfields of AI research. But—as this survey suggests—it is important to keep in mind that these subfields are not necessarily reflective of AI researchers as a whole.

Methodology

We designed a survey to ask AI researchers about their compute use, perspectives on the role of compute and other resources in their research and in broader AI progress, and opinions on government-provided compute resources. We surveyed AI researchers on this topic for several reasons. First, there is no comprehensive data available on compute use among AI researchers, and attempts to measure use focus on compute-intensive models, which might not reflect the broader AI research community.[11] Second, while the past decade has seen dramatic increases in compute use, it is unclear whether or how insufficient compute access impedes AI progress. A survey allowed us to assess compute use among a broader set of researchers and to ask questions specific to the role of compute in driving or impeding research progress.

We define AI researchers as individuals who have authored a paper in a top AI conference or journal, or who work in industry in an AI-related role. This sampling frame is in line with other recent surveys of this population, which use AI conference participation, research publication, job titles, or AI-relevant skills as criteria for inclusion (see Appendix B). We identified authors of papers in 20 leading AI journals or conferences between 2016 and 2021 using Web of Science (see Appendix A for the list of AI journals and conferences).[12] This resulted in 27,172 authors with email contact information who were affiliated with a U.S. institution at the time of their paper's publication. Second, we identified industry AI researchers using LinkedIn data from Revelio Labs.* We looked for LinkedIn users who listed (1) their job as a machine learning or artificial intelligence engineer (or similar); or (2) their employer as one of 46 AI startups and their job as a technical role (see Appendix A for included job titles).† We randomly selected roughly five thousand profiles that met this criteria and used RocketReach, an email sourcing vendor, and manual searching to identify emails for 3,894 industry AI researchers.[13]

---

* LinkedIn data is provided to CSET by Revelio Labs, a workforce intelligence company (reveliolabs.com). Note that the sampling method used here is not likely to include AI engineers in industry who are focused on deploying large systems at scale. While authors who publish in top AI journals or conferences likely include a number of employees at large tech companies, those respondents probably focus on research as opposed to deployment at scale of AI systems.

† Our list of 46 AI startups was taken from "The United States of Artificial Intelligence Startups," *CB Insights*, August 4, 2021, https://www.cbinsights.com/research/artificial-intelligence-startup-us-map/. We wanted to include researchers from AI startups because they fall within the sampling frame used here, but they publish less research in top AI venues, so may not be captured by that sampling method. Including this list of AI startups also encouraged geographic diversity in our sample.

In total, we received 410 complete responses (and 123 partially completed responses, which were also included in the analysis), for a response rate of 1.7 percent.[14] For a comparison of our response rate to those achieved by similar surveys, as well as an analysis of nonresponse bias in our results, see Appendix B. The median survey response time was eight minutes.

The survey included 30–35 close-ended questions and one open-ended question, based on respondents' reported employment experiences and AI projects. Respondents were asked about their AI projects, compute usage, research priorities, and opinions regarding the importance of various factors for AI research progress. Early versions of the survey instrument were refined through a series of cognitive interviews with AI researchers in academia and industry.[15] The full survey instrument is available at the project GitHub repository.[*] Since response rates for online surveys tend to be low compared with other modes of distribution, we carefully considered ways to boost response rate and chose to ensure anonymity, clearly articulate the research goals in the invitation email, and distribute two reminders after the initial survey distribution. No compensation was offered for survey participation.

---

[*] The GitHub repository is available at https://github.com/georgetown-cset/Compute_Survey_2022/.

Results

Respondent demographics

Of our 410 complete responses, 275 (67 percent) reported working in academia, 120 (29 percent) in industry, and 14 (3 percent) in government.[*] Among respondents who reported working in industry, 84 reported working for a company with more than 500 employees, while 35 reported working for a company with 500 or fewer employees (see Figure 1).

Figure 1. Survey Respondent Employment Sector
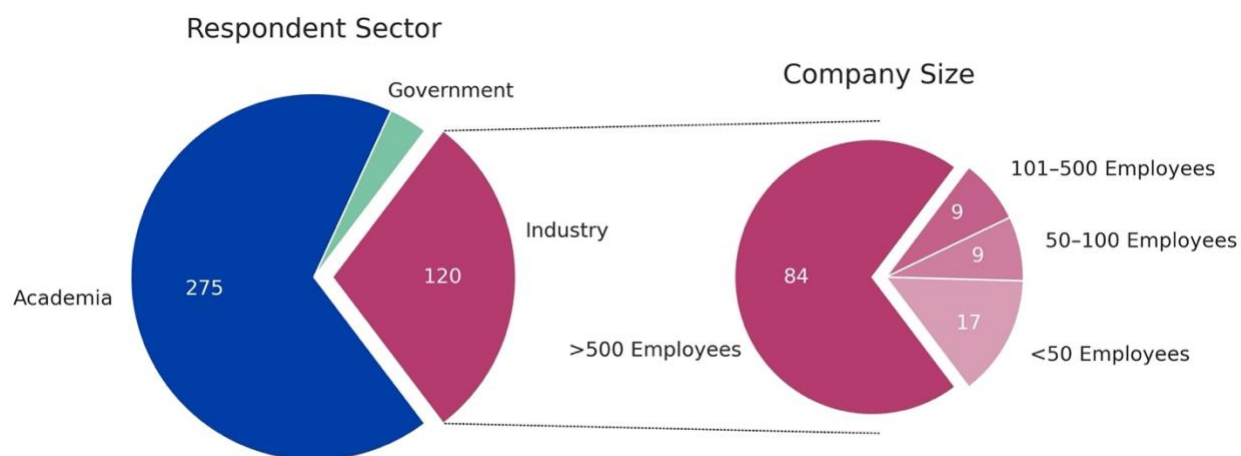
Source: CSET Compute Resource Survey

To help understand this sample of academic respondents, we looked at the email domains for all AI researchers invited to participate in the survey who started or completed it.[†] That set included 423 ".edu" email domains: 147 (35 percent) from a top

---

[*] One respondent selected affiliation as "None of these," while another also indicated working in industry but did not respond to the question about organization size. No respondent who only partially finished the survey indicated a sector affiliation. See Appendix B for an overview of the representativeness of this breakdown as compared to the larger population invited to participate in this survey. We did not make explicit efforts to include government respondents in the sampling frame, as existing policy discussions primarily focus on perceived differences in compute access between industry and academia.

[†] Due to anonymous response collection, we cannot match specific emails to responses or get the distribution for only the responses included in the analysis. In Appendix B, Table B.2, the number 634 includes respondents who "finished" the survey in the sense that they were screened out or did not consent, at which point the survey ended.

50 university, 115 (27 percent) from a university ranked 51–200, and 134 (32 percent) from a university ranked below 200, according to QS World University Rankings.[*] This suggests our sample includes researchers working in different tiers of academic institutions.

Respondents were also asked to indicate which AI fields they worked in (with top-level options consisting of computer vision, natural language processing, reinforcement learning, robotics, or other); the number of respondents that reported working in each field is shown in Figure 2.

Figure 2. Survey Respondent Reported AI Fields
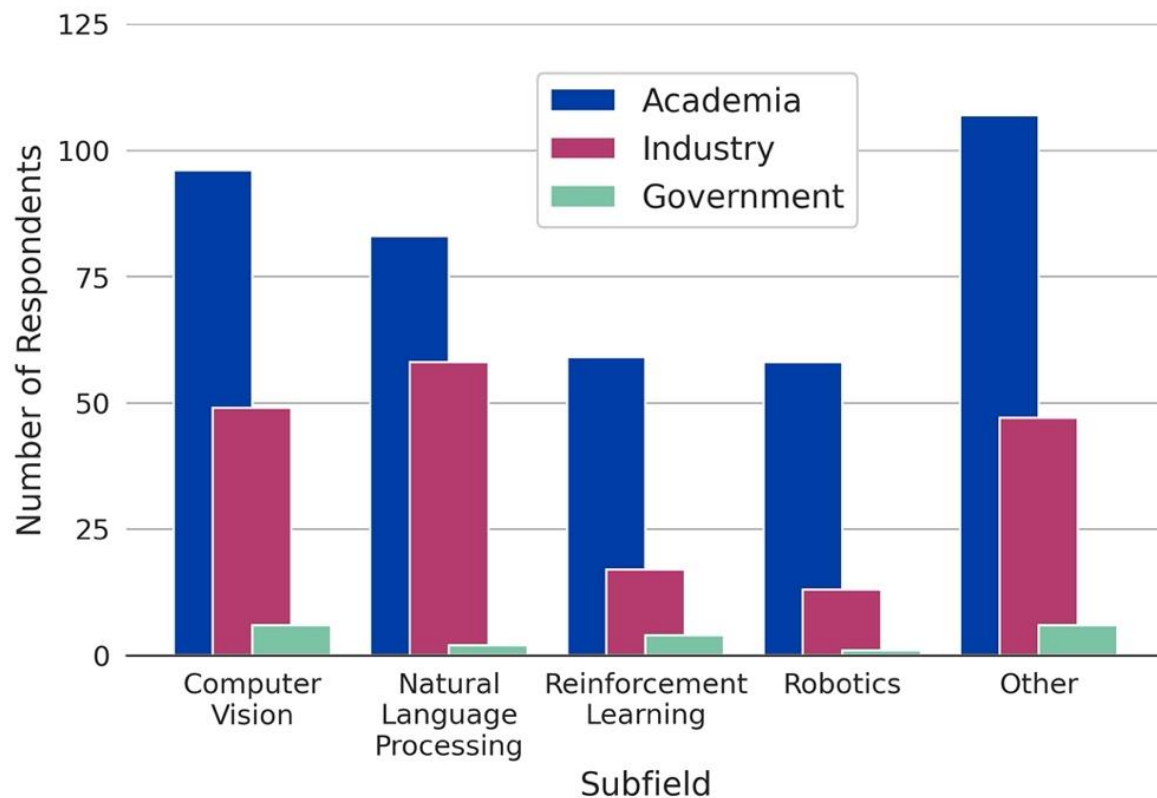
Source: CSET Compute Resource Survey

Comparing reported AI fields for academic- and industry-affiliated respondents, a larger share of academics reported working in robotics and reinforcement learning, while among industry respondents, a large share reported working in natural language

---

[*] The remaining 6 percent were associated with universities not ranked by QS World University Rankings.

processing.* Respondents could also indicate specific subfields they worked in (e.g., object tracking). Full breakdowns of the number of respondents by field, subfield, and sector can be found in the GitHub repository, and Appendix C contains comparisons between subfields across each of the five top-level categories.

Compute is not the primary constraint for many AI researchers

A goal of the survey was to understand how AI researchers see compute as a resource driving or constraining their research. We included several questions to capture these perspectives, including asking researchers to report the relative importance of compute, data, and talent for their projects. We also asked what resources they would prioritize given a larger budget, how often compute and other resources caused them to abandon or revise a project, and the importance of compute in driving AI progress to date and in the future.

Finding 1.1. Researchers report talent as the primary factor contributing to the success of their most significant projects.

Respondents were asked to share details about two projects they worked on in the previous five years: the project that they felt made the most significant contribution to research progress in their field ("most significant project"), and their most compute-intensive project. Interestingly, 67 percent of respondents reported that these two projects were the same.

While most surveyed researchers viewed their most compute-intensive project as their most significant project, they rated other factors as more important to the project's success. Asked directly how important various factors were for their most significant project, 90 percent rated "specialized knowledge, talent, or skills," and 52 percent rated "large amounts of compute" as very or extremely important for the same project's success, as shown in Figure 3.[16]

Figure 3. Percent of Respondents Viewing Factors as Important for Project Success
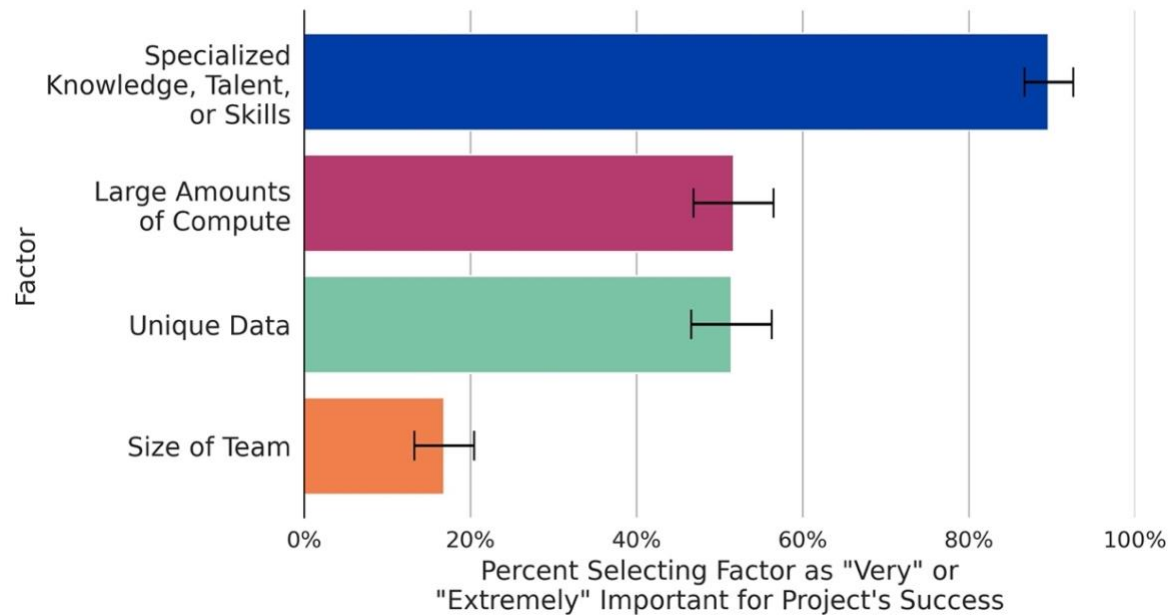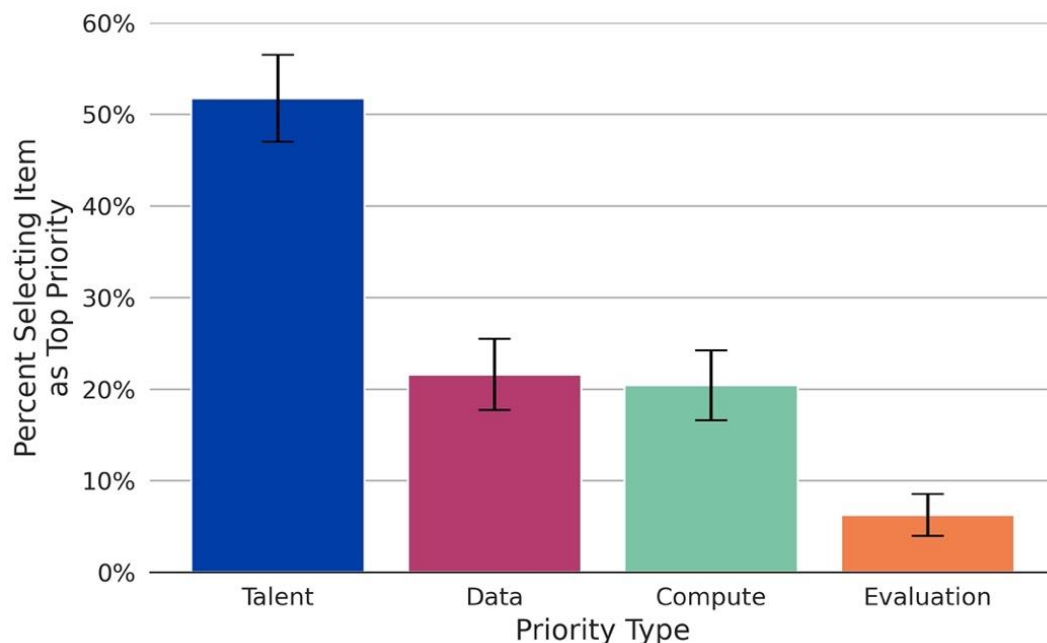

Source: CSET Compute Resource Survey
Note: Bars represent 95% confidence intervals.

A similar proportion (51 percent) rated "unique data" as very or extremely important. This question asked respondents to rate each factor independently, but other questions asked respondents to compare compute with other factors, and talent again surfaced as an important resource.

Finding 1.2. Most researchers would prioritize talent if they had more funding.

Managing an AI project is, for many researchers, a matter of carefully overseeing a project budget, which must be stretched to cover salaries, data collection, compute costs, and testing prior to deployment or publication. To assess how researchers prioritize compute when allocating their budget, we asked them to imagine the budget for their current or most recent AI project doubled: What would their first priority be to spend the money on?

Figure 4. Percent of Respondents Selecting Factors as Their Top Budget Priority

Source: CSET Compute Resource Survey
Note: Bars represent 95% confidence intervals.

Roughly half (52 percent) said that they would first spend the additional money on either "hiring researchers" or "hiring more programmers or engineers," which are binned together in Figure 4 under "Talent."* About a fifth of researchers would make "purchasing more or higher-quality compute" their first priority, and a similar share would first use the funds to collect or clean data.

This question does not capture the actual budget amount spent for any of these categories. It is possible that when starting a new project, researchers prioritize funds for compute, only later finding that they would like more money to spend on talent. At the same time, compute is generally more fungible than data or researcher access. It would generally be easier to convert extra funds into more compute than to use them to get better data or a larger research team. Our finding that most respondents would still choose to use additional funding to hire more people, regardless of allocation of any existing budget across these resources, suggests talent may be a more pressing concern for researchers, relative to compute, over the full project life cycle.

---

* Two other choices—"collecting more data" and "refining or cleaning data"—were binned under the heading "Data."

Finding 1.3. When researchers are forced to change their research plans, it is more often due to talent or data limitations than to compute limitations.

One indication that a factor is a constraint on progress is that researchers frequently change their research plans due to insufficient access to that factor. To explore this possibility, we asked researchers how often, over the past two years, they (1) rejected a project; (2) revised an ongoing project; or (3) abandoned an ongoing project due to (a) insufficient compute; (b) insufficient data; or (c) insufficient researcher availability.[*] Responses were recorded as one of five options, ranging from "never" to "all the time," and the mean responses for each question are displayed in Figure 5.

Figure 5. Rates at Which Respondents Change Research Plans Due to Various Factors
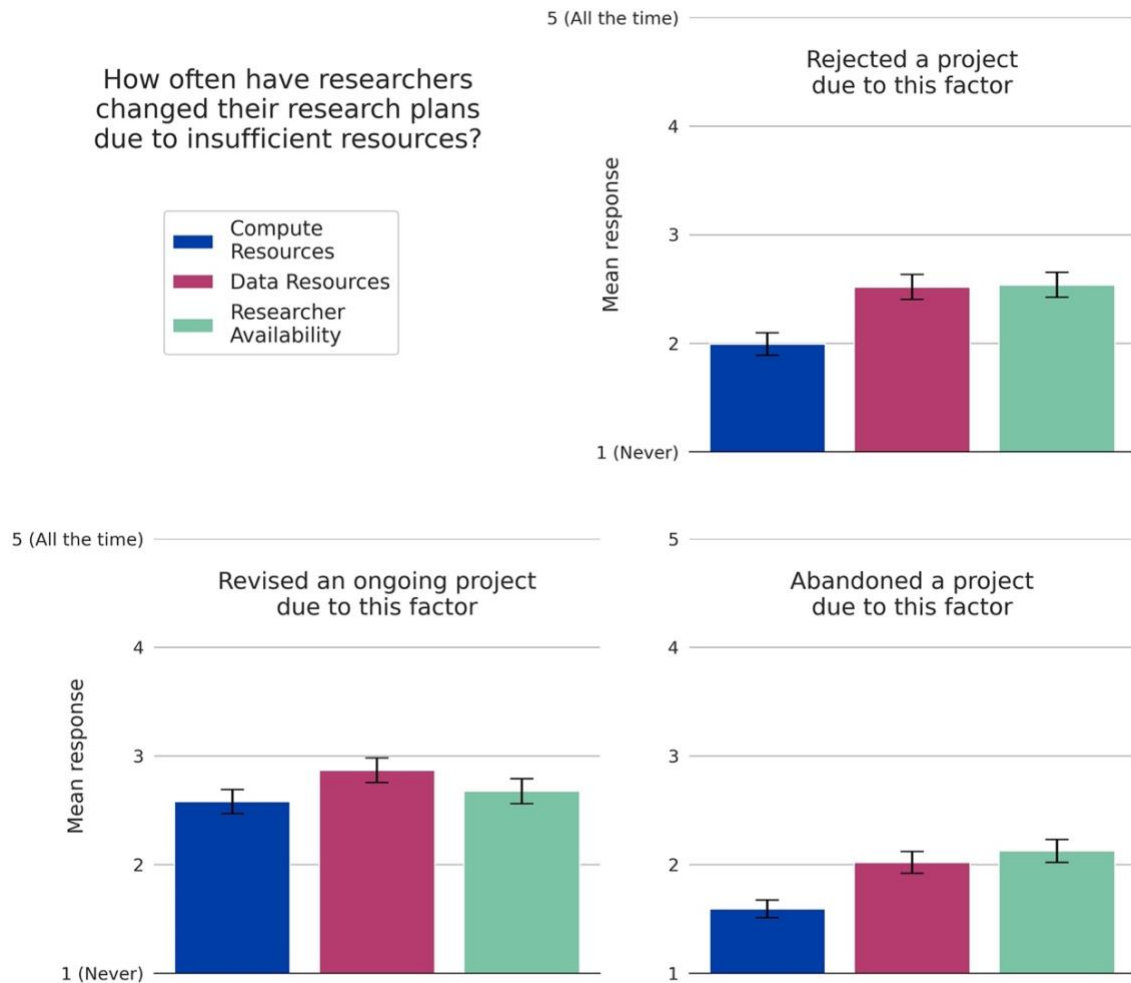
Source: CSET Compute Resource Survey
Note: Bars represent 95% confidence intervals.

---

[*] We randomized the order of AI resources presented to respondents.

Researchers report rejecting and abandoning projects due to a lack of data or researcher availability more often than due to a lack of compute resources. In addition, a lack of data (but not of researcher availability) is more often reported as the reason for revising ongoing projects than a lack of compute resources.[17]

While lack of data and talent is more often given as the reason for rejecting or abandoning a project, 76 percent of respondents report revising projects due to insufficient compute at least sometimes during the past two years. This result is consistent with a recent survey conducted by the Organisation for Economic Co-operation and Development Expert Group on Compute and Climate, which found that a similar proportion of respondents reported challenges in accessing sufficient compute.[18] The OECD survey did not ask respondents whether they also faced difficulties accessing data or talent. The OECD concluded from its survey that compute is currently receiving insufficient attention from policymakers relative to these other factors, but our findings add nuance to this argument: while compute does constrain AI researcher projects, data and talent do so more frequently.[19]

Our findings do not capture the possible case of researchers not even considering projects because they know in advance that they will not have sufficient resources. When developing the survey, we were interested in exploring whether researchers think about hypothetical projects, such as training a language model to rival GPT-4—the basis for ChatGPT—but reject them knowing they do not have the necessary resources.[20] After testing some questions to study this possibility, we decided that we could not reliably measure rejected hypothetical projects through survey measures. This means our findings cannot speak to cases in which researchers may not even consider a project due to lack of compute, data, talent, or other resources. We did, however, find that 43 percent of respondents reported never rejecting a project due to insufficient compute, which indicates that some subset of AI researchers are able to pursue the research they want at their current level of compute resourcing.[*]

Finding 1.4. Most respondents think computing's role in driving AI progress will stay the same or decrease in the next decade, compared to its role in the past decade.

We asked respondents for their level of agreement with the claim that progress in AI over the *past decade* was the result of five different factors: data, compute, algorithms,

---

[*]While academic respondents more frequently report rejecting a project due to a lack of compute when compared to industry researchers ($p = 0.029$ by a Mann-Whitney U test), 39 percent of academics say they never reject projects due to insufficient compute.

number of researchers, and level of support for AI projects. Each statement read: "Progress in AI over the past decade was the result of [factor]." There was general agreement that each factor contributed to AI progress during this period, with 59 percent indicating strong agreement that past AI progress was the result of more compute—higher agreement than any other factor.

While surveyed researchers agreed that increased compute was critical for AI progress to date, fewer respondents strongly agreed that it would be a driver of AI progress *over the next decade*. Compared to 59 percent of respondents who indicated strong agreement that more compute was a past driver of AI progress, fewer (40 percent) strongly agreed that more compute would drive future AI progress. One factor increased in strong agreement among respondents—better algorithms. Specifically, 31 percent strongly agreed it was a driver of past AI progress, but 53 percent strongly agreed that it would drive future progress—the highest jump in agreement among the factors. Table 1 shows the change in strong agreement for each factor's influence on past and future AI progress.

Table 1. Respondent Views on the Importance of Various Factors for AI Progress

Factor	Percent strongly agreeing that AI progress was/will be the result of this factor over the:		Change (in percentage points)
	past decade	next decade	
Better algorithms	31%	53%	22% ***
Greater support for AI projects	33%	42%	9%
More researchers in the field	32%	35%	3%
More or better data	47%	44%	−3%
More compute	59%	40%	−19% ***

Source: CSET Compute Resource Survey

Note: Asterisks indicate statistically significant differences (p < 0.001) as calculated by a Mann-Whitney U test with Bonferroni correction comparing past decade to next decade responses.

This result warrants some discussion. Predictions that the importance of algorithms will rise while the importance of compute falls could be a reflection of the researchers' own interests rather than a developing trend. Researchers tend to view progress that comes from the brute force approach of simply using more compute as less interesting

and valuable than new approaches that require their knowledge and creativity. That more compute often outperforms more ingenuity has been a "bitter lesson" that has perhaps still not been entirely internalized by the research community.[21]

At the same time, while increases in compute power have pushed AI dramatically far forward in the past decade, there are reasons to suspect that the past decade's trendline of skyrocketing compute usage cannot be sustained.[22] If this is true, it is reasonable to expect—as two of the present authors have argued—that future AI progress will rely increasingly more on algorithmic improvements compared to the past decade of AI research. This interpretation is also consistent with our survey results, because the results displayed in Table 1 may demonstrate that a meaningful number of AI researchers have arrived at similar conclusions about the future of research in their field.

Reported compute use is similar for industry and academia

Another goal of the survey was to examine differences between academic and industry researchers in compute use and needs. In this section, we break down responses to various questions included in the survey intended to capture compute use and access according to the respondent's reported employment in academia or industry.[*]

Finding 2.1. Academics report paying less for compute but do not report significantly less compute use.

Respondents were asked several questions about the most compute-intensive AI project they had worked on in the preceding five years.[†] When asked how expensive the total compute required by this project was, academics reported spending significantly less than industry researchers, as shown in Figure 6. This finding is consistent with the narrative that the compute capabilities of industry researchers are rapidly outpacing those of their academic counterparts. When asked about compute use for this same project in terms of GPU hours, however, we observe no meaningful difference, also shown in Figure 6.[23]

---

[*] We omit government researchers due to small sample size.

[†] We also asked a similar set of questions for their most significant project. To compare the level of compute use between industry and academia, we focus on researchers' most compute-intensive projects, as our aim here is to better understand the variability in researchers' maximum compute access and need.

Figure 6. Reported Compute Use in Cost and GPU Hours for Respondents' Most Compute-Intensive Projects
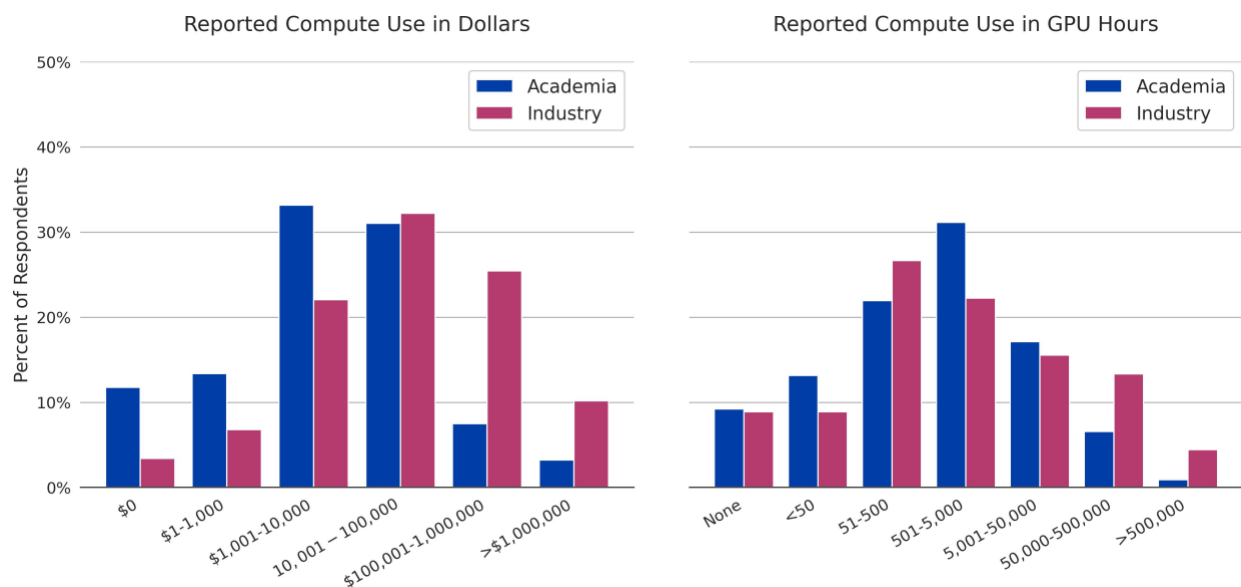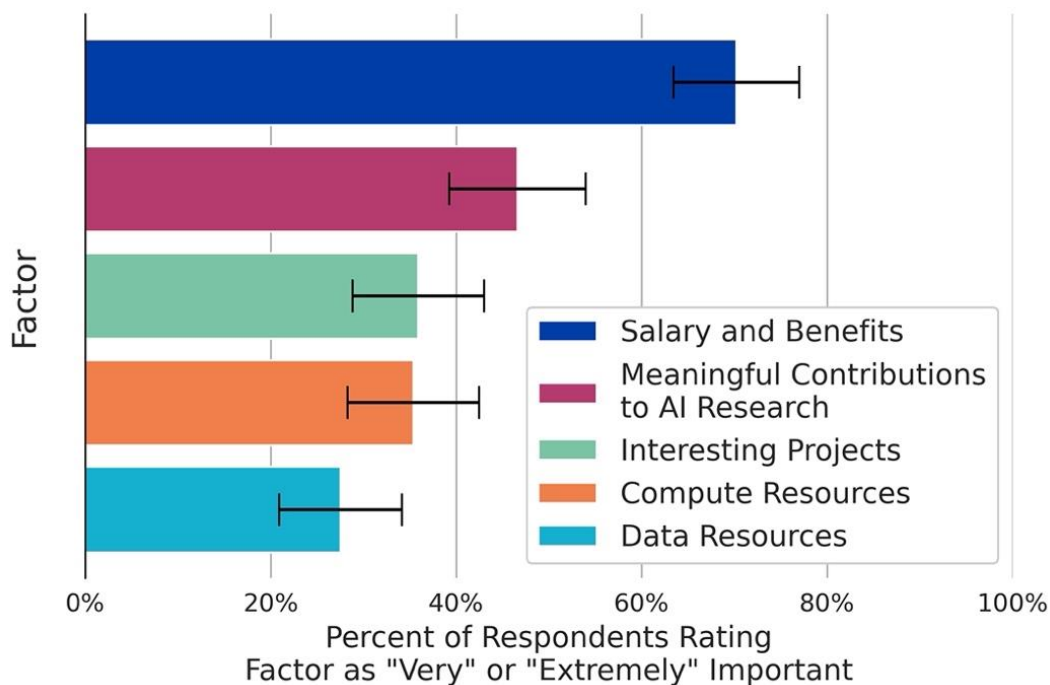

Source: CSET Compute Resource Survey

While we find no reported difference in compute use, as measured by GPU hours for respondents' most compute-intensive project, we acknowledge this does not capture all possible differences in compute access between industry and academic researchers.[24] We nonetheless regard GPU hours as the better measure for compute use for several reasons. First, 349 respondents provided information about GPU hours, compared to only 278 respondents for cost.[25] Second, it may be that researchers who use on-premise compute—which has already been paid for—report "$0," and more on-premise users are academics.[26] Third, cloud computing companies often provide access to compute resources at discounted rates for academics. Combined, these factors make monetary cost a less reliable measure of compute use across sectors.

Finding 2.2. Academics cite salary and benefits as an important consideration for leaving academia more often than compute resources.

When asked if they ever considered leaving academia for an AI-related role in industry, 65 percent of academic respondents answered yes, underscoring the risk of universities losing researchers to private industry.[27] Among academics who answered yes, 70 percent cited salary and/or benefits as a very or extremely important factor in considering leaving academia, as shown in Figure 7.

Figure 7. Importance of Various Factors to Academics Who Have Considered Leaving Academia
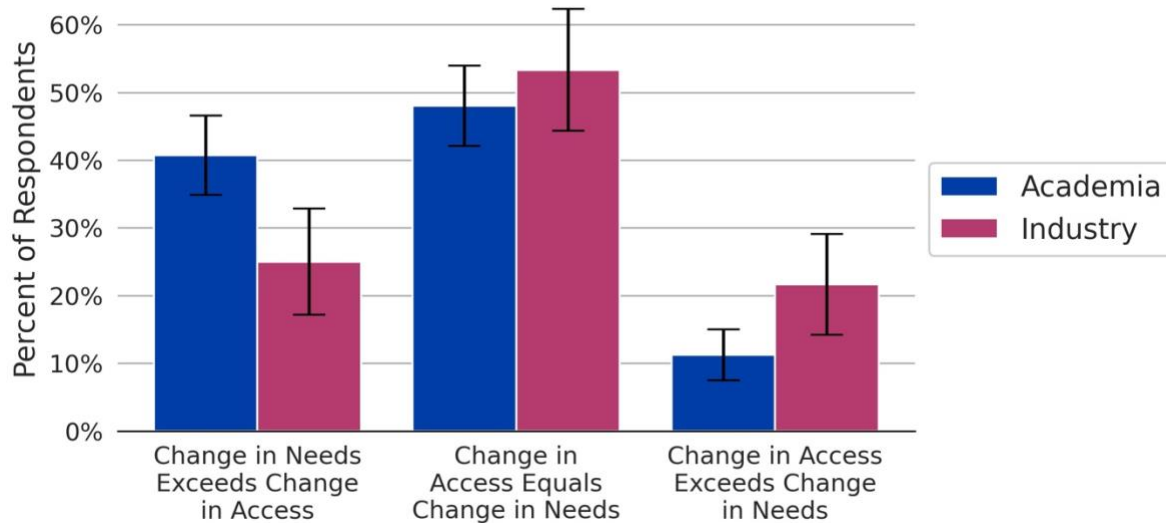

Source: CSET Compute Resource Survey
Note: Bars represent 95% confidence intervals.

The factors least often rated as very or extremely important were compute or data resources, with 35 percent and 28 percent of researchers rating them as very or extremely important, respectively. This is consistent with prior CSET survey research, which found data and compute resources to be the least important consideration for AI PhD graduates in deciding where to work after graduation.[28]

Finding 2.3. Academics report that compute needs have outpaced availability, but they are not significantly more concerned about future access impacting their contributions to AI.

We also asked respondents how much compute they *need*, relative to two years ago, and how much compute they *have access to*, relative to two years ago. The results are shown in Figure 8. We observe a significantly greater proportion of respondents in academia reporting that their change in compute needs has exceeded their change in compute access, as compared to respondents in industry.[29] This suggests that academic research is likely to be increasingly constrained, by comparison with industry research, as compute needs increase.

Figure 8. Responses Regarding Changes in Compute Access and Compute Needs by Sector

Source: CSET Compute Resource Survey
Note: Bars represent 95% confidence intervals.

We also asked respondents to what extent they were concerned that a lack of compute resources would be an obstacle to their contributions to AI in the next decade. Figure 9 compares responses for academics and industry researchers, and reveals little difference in level of concern. Academics were slightly more likely to report being "moderately" or "extremely" concerned, but those differences are not significant.[30]

Figure 9. Respondent Concern over Future Compute Access by Sector
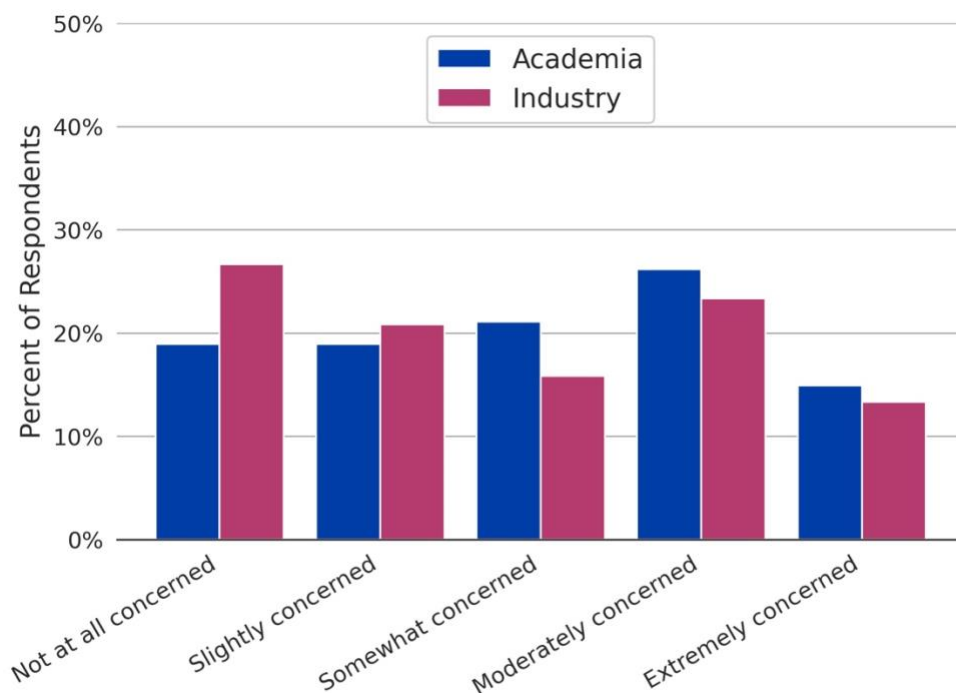
Source: CSET Compute Resource Survey

Returning to the common narrative that AI researchers in academia have less access to compute and greater concerns about their level of access relative to industry researchers, we find some support for a growing gap in access, but no support for a higher level of concern among academics. In terms of differential access, we also look at whether researchers with the least compute—regardless of their affiliation—are most eager to receive more.

Finding 2.4. Higher compute use correlates with being more concerned about compute.

Examining the relationship between current compute use and future concerns, we observe some interesting trends. Figure 10 displays respondents' mean level of concern about having insufficient compute to contribute meaningfully to AI research in the future, according to respondents' reported GPU hours for their most compute-intensive project.[*] This figure shows that on average, respondents who report using

---

[*] Note that the two lowest responses (no GPU hours and <50 GPU hours) and the two highest responses (50,001–500,000 GPU hours and more than 500,000 GPU hours) are combined in this figure due to small sample sizes at the extreme ends of the range for GPU hours.

higher amounts of compute express more concern about having sufficient compute to contribute to research in the future, though the differences between academics and industry researchers at each level are not statistically significant.[31] The group that reports being most concerned about insufficient future compute access is academics at the upper end of the compute use range.

Figure 10. Mean Level of Concern over Future Compute by Compute Use and Sector
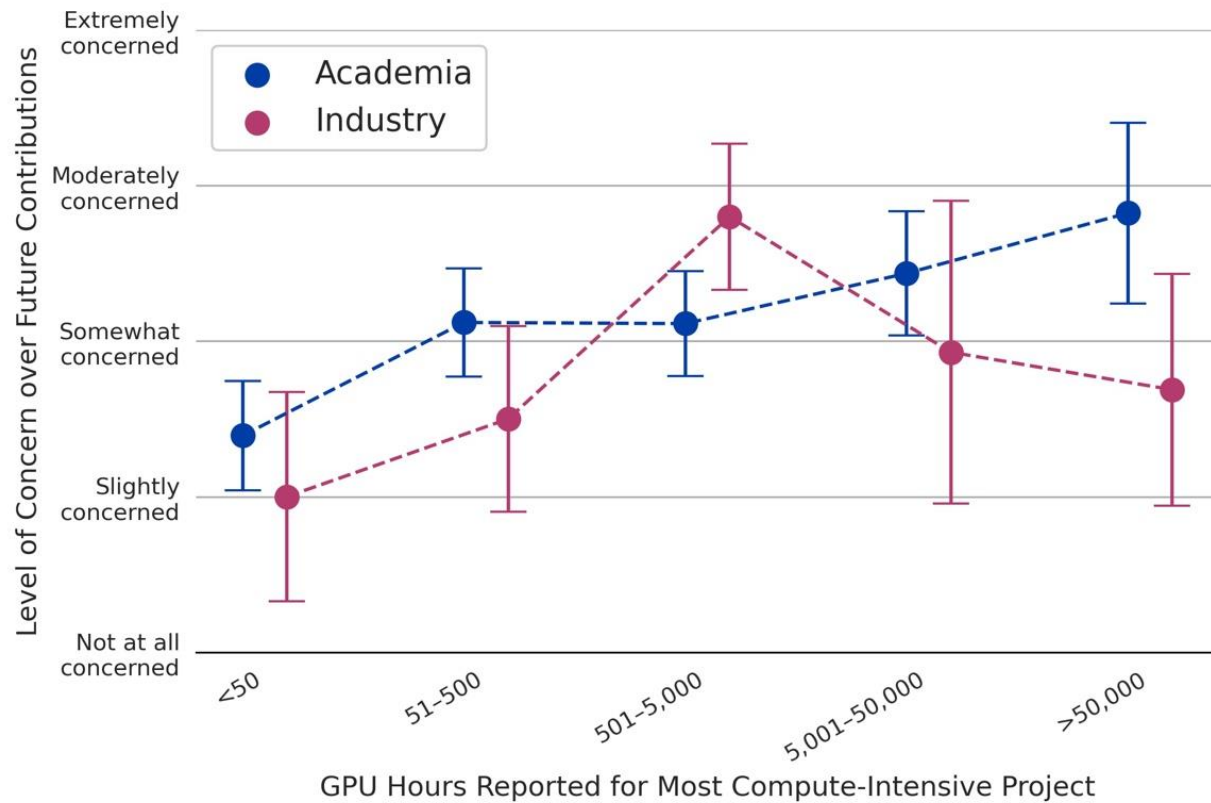
Source: CSET Compute Resource Survey
Note: Bars represent 95% confidence intervals.

We also revisit other survey questions to see if any differences exist based on reported compute use. We find that higher reported compute use is positively correlated with each of the following: more frequently changing project plans due to a lack of compute, considering compute an important factor in leaving academia, and agreeing that compute has been a driver of AI progress over the past decade and will continue to be over the next decade, as shown in Box 1.

> **Box 1: Correlation between reported compute usage and compute attitudes and behaviors**
>
> Respondents who reported using greater amounts of compute in their most compute-intensive project also tended to give greater responses on each of the following indicators:[*]
>
> - Frequency of abandoning a project due to insufficient compute ($\rho = 0.11$)
> - Frequency of revising a project due to insufficient compute ($\rho = 0.27$)
> - Frequency of rejecting a project due to insufficient compute ($\rho = 0.28$)
> - Importance assigned to a lack of compute as a reason to consider leaving academia ($\rho = 0.33$)
> - Level of agreement that compute was a major driver of AI progress over the past decade ($\rho = 0.19$)
> - Level of agreement that compute will be a major driver of AI progress over the next decade ($\rho = 0.17$)
>
> Source: CSET Compute Resource Survey

One explanation might be that researchers' current level of compute use is influenced by self-selection: Researchers choose to pursue work in more computationally intensive subfields or to adopt particularly computationally intensive research methods. Self-selection into these fields and methods could then shape levels of concern and the need to revise research based on compute access. But this might also mean that researchers who already use a lot of compute would be the most motivated to seek out and make use of new compute resources. In this case, attempts to provide more compute to researchers broadly could increase any existing divides between high and low compute users.[†]

---

[*] We report Spearman's rank correlation. The p-values for the six indicators discussed here were 0.045, <0.001, <0.001, <0.001, 0.001, and 0.003, respectively. These values were calculated based on a permutation test with 10,000 samples. The correlation between compute use and importance assigned to a lack of compute as a reason to consider leaving academia includes only responses from academics; all other correlations include respondents from all sectors.

[†] In addition to helping some researchers more than others, subsidized access to large amounts of compute may help computationally intensive branches of study—such as foundation models—more than others. These branches do not represent the majority of our respondents, and they may or may not be deserving of extra policy support. If self-selection effects help to explain why some AI researchers use more compute than others, then further research should focus on analyzing how these self-selection effects operate, because such research would help to identify how policy interventions could more effectively democratize compute usage.

Researchers have a variety of opinions about a national AI research resource

To evaluate what support researchers would most like from the federal government, we asked respondents to select among five different resources they would find useful for the government to provide. The results, broken down by the respondent's sector, are shown in Figure 11. The survey question did not mention specific proposals such as the NAIRR, but told respondents that "the U.S. government is exploring creating a national AI research resource, which could provide various types of support to AI researchers." The two follow-up questions were designed to elicit responses that could inform policy discussions around the current formulation of such a resource within the U.S. government.[*]

Figure 11. Desired Government-Provided National AI Research Resources by Respondent Sector
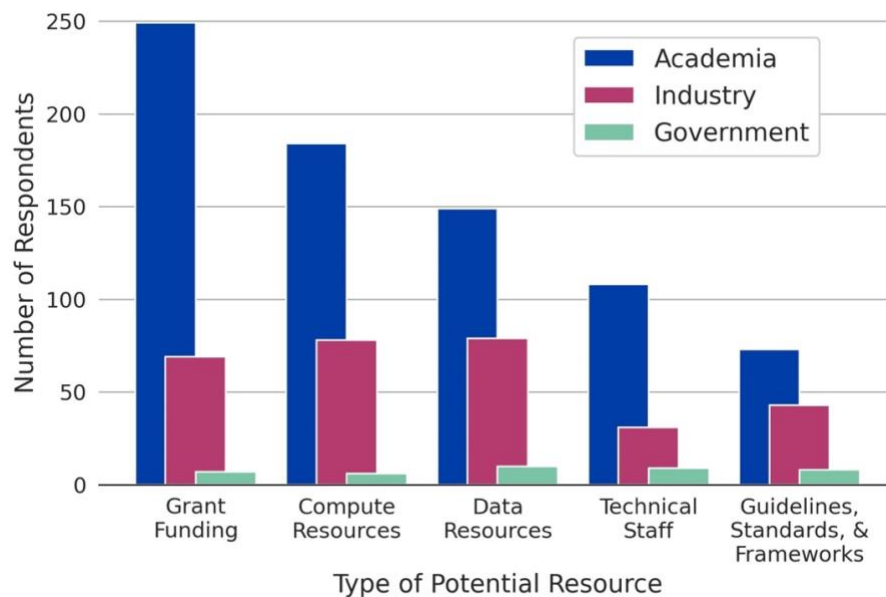
Source: CSET Compute Resource Survey

Overall, grant funding was the most frequently selected resource across all respondents. However, more respondents indicated interest in government-provided computing resources than in data or technical staff.[32] One possible explanation for this interest in government-provided compute resources, despite our finding that researchers are more often constrained in practice by data and talent resources, is that

---

[*] The title of this report is drawn from one of our responses to the open-ended question, "Do you have any other thoughts or suggestions about the U.S. government creating national AI research resources?"

researchers view resources that afford them greater agency as more helpful. For example, grant money can be used for a wide variety of purposes, but compute resources must be expended on actual computation. While technical support staff may be broadly useful, respondents may have interpreted this category in different ways or may not trust talent that they have not personally vetted to work on their projects. A smaller number of respondents would find "guidelines, standards, and frameworks" useful, though there may have been some ambiguity in how respondents interpreted this phrase.

We invited respondents to share any thoughts or suggestions they had for the U.S. government as it works toward creating national AI research resources. Eighty-five of our respondents provided an answer to this open-ended question. Responses focused on the five broad categories asked about in the preceding question—compute, data, staffing and workforce development, grant funding, and standards and frameworks—but many responses also offered broader suggestions about the implementation of government-led initiatives to support AI researchers. Themes for each of these categories are discussed next, as they present an opportunity to qualitatively assess the priorities of our respondents.[33]

Compute Resources

When asked to share their thoughts on the implementation of government-provided AI research resources, 38 of the 85 respondents specifically commented on compute. A majority of these responses addressed the potential utility of compute resources to their own research or to the AI field as a whole. For instance, one respondent noted that "papers are often rejected on the grounds of limited experiments, which were in fact limited because of a lack of compute, not because of lack of interest or researcher time."

"I'd love to have compute resources, but I'm intentionally choosing the projects where I'm not blocked by these things—and there are several lifetimes worth of projects in these areas."

Many of these responses specifically highlighted perceived gaps between academia and industry, for instance that "the amount of compute resources . . . available to a typical academic/university research lab is multiple orders of magnitude smaller than what is available to researchers at large tech companies." This gap was noted as a cause for concern by multiple respondents, with one respondent also mentioning a worrying

gap in resources between startups and larger tech firms. While it is likely that "large tech companies" have more access to compute than does a typical academic lab, with this survey we were unable to substantiate claims that academics systematically have access to less compute for individual projects than do industry researchers.

However, eight respondents expressed skepticism about a compute-heavy approach to designing national AI research resources. In several cases, respondents indicated that they believed the survey instrument itself improperly confounded AI research as a whole with computationally-heavy deep learning research. Other respondents suggested that compute access might not enable as much research progress as is commonly believed, or that it would fail to address existing research gaps. Comments along these lines included those from respondents who stated that "there's a ton of research that could be done on much smaller scales," or that "not many efforts have been spent in understanding why . . . progress has been driven by computationally heavy research." Finally, five respondents expressed skepticism that government-provided compute resources would be sufficiently accessible, well-supported, and cutting-edge to be sustainably useful even as hardware technology improves.

Data Resources

Seventeen respondents noted data and data accessibility, including the need for open, accessible datasets and related resources (e.g., source code, repositories, models, etc.) to advance AI. Multiple respondents noted the need for diverse and user-friendly data sets, with one observing that "there is a great need for real-world, sensitive data . . . We need better privacy laws in general, but we also need much more personal human data available to researchers to make progress toward human-centered technologies." While a relatively small number of respondents specifically mentioned data issues in their answers, most of these answers emphasized the difficulties around accessing well-cleaned, curated, and maintained datasets.

Talent, Staffing, and Workforce Development

Twenty-three respondents discussed talent- and workforce-related resource development. Many of their responses called for policy changes or emphasized structural problems in these areas. One emphasized that the AI workforce is limited in part by the difficulty that foreign PhD students face in acquiring green cards: "Foreign students complete amazing work on their PhDs and then struggle to continue after graduation, having to settle for job offers that help them stay instead of work that is relevant to the field." Another respondent emphasized that the movement of academics into private industry creates risks of "state-of-the-art AI technology

[becoming] monopolized and controlled by a small number of corporate entities, [which] is extremely risky in terms of both its economic and national security implications."

Some respondents specified that workforce development goals should focus on supporting researchers in specific fields. For instance, in the words of one: "By far the most important national AI resource need is a safety/correctness brain trust staffed by researchers who have post-PhD industry experience building AI systems." In addition, several respondents who emphasized the workforce were explicit that they saw its development as a higher priority than compute: "I can take a CS undergrad and teach them what they need to know in order to be helpful in my lab in terms of coding, and they can probably also pick up specific algorithms/techniques from reading papers. That part is easy. Having somebody who can help design and think through a data collection and labeling process, and who understands how to work with stakeholders and domain experts to bridge the communication and knowledge gap with ML researchers—that's the hard part."

"I want my tax dollars to fund 30–40 year investments in PEOPLE, not 3–5 year investments in hardware."

A few respondents noted a need for supplemental technical staff support when discussing their concerns about the accessibility of national AI research resources. This need was rooted in the concern that the compute resources themselves would, while potentially providing value, have a learning and transition cost: "compute resources are not standardized enough at this point when it comes to AI; so it is difficult for PhD students and junior researchers to directly switch to using government-provided compute resources; *unless* the said resources also come with technical staff who can manage the transition and guide development."

Guidelines, Standards, and Frameworks

Twenty-seven respondents commented on some topic that related to the broader subjects of guidelines, standards, and frameworks, typically mimicking at least one of those terms from the preceding question, but in many cases appearing to interpret the terms differently. Some respondents who emphasized the need for "frameworks" seemed to interpret this as meaning specific technical resources such as "open source and pre-trained GPT3 and DALL-E or more stable TensorFlow or ROS [Robot Operating System]." By contrast, others asked for legal or ethical guidelines—a meaningfully distinct type of government resource. While we deliberately tried to

leave this category broad, it is possible that the level of ambiguity made respondents hesitant to select that option in the preceding question.

Among the respondents who spoke about the need for guidelines, standards, or frameworks, 12 seemed to interpret this as meaning technical standards, open-source tools, or evaluation metrics. One requested "evaluation resources: testbeds that can evaluate interactive AI systems with a diverse pool of human users in realistic settings," while another asked for "specific ambitious challenges, with well-defined metrics." A third respondent spoke of lacking resources in academia to close the research gap between academia and industry, encouraging the federal government to provide "open source code and software libraries like Theano,[*] Tensorflow, and Pytorch [which] made things progress fast."

"I am also very glad to see 'guidelines, standards, and frameworks' listed there, as that is really important—again, scale up NIST and help them get their message, skills, and tools out into the world."

Another eight respondents discussed the need for ethical or legal frameworks for developing, using, and evaluating AI systems. One respondent expressed a desire to see the government "enforce standards around *ethics* in AI—in the collection and usage of data, in transparent AI, in model monitoring and evaluation, and in use-case applications." Other respondents noted a need for legal guidelines for AI development and use, with one stating that "the government is responsible for regulating and setting standards for AI, like everything else. It should have an agency responsible for adapting human rules to AI . . . It is not fair to the researchers to get prosecuted after the fact for something there were no laws on. And it is dangerous to let companies take the research in any direction they like, which may harm people."

---

[*] Theano was an early open-source machine learning library maintained by researchers at the University of Montreal. In 2017, the institute responsible for maintaining Theano announced that it would stop developing it due to the proliferation of alternative frameworks.

Grant Funding

Twenty-five respondents indicated a preference for grant funding and offered specific guidance on what types of people or projects those grants should provide resources to support. A typical comment in this regard was that "grant funding would be more useful than compute resources." Nine respondents encouraged further research or investment into a particular subfield of computer science. Specific areas included "research activities that explore 'small-data' algorithms that may have better utilization across the world" and "AI research that produces public goods like the prevention of catastrophes."

"Government is generally bad at predicting what resources will be needed. I think it's better to give funding and let the users themselves determine what they need and allocate resources accordingly."

General Suggestions

General suggestions were popular with respondents; 51 offered comments that did not directly address any of the topics previously discussed. These responses touched on what government-led AI research resource provision should focus on, how it should allocate funding, how it should be limited, how its resources should be distributed, and similar areas. Within this broad category, one of the most consistent points was the importance of reaching out to a diverse set of researchers and supporting AI research in an inclusive and equitable way. One respondent noted a need "to address bias, diversity, ethics and inclusion of these systems," and another stated: "It is not more researchers necessarily [that we need], but more diversity of researchers could help advance the field." Multiple respondents explicitly stated that government-provided resources should provide "support to a wide variety of individuals/researchers in place of just supporting the big or known personnel."

Beyond the issue of equity and inclusion, a few respondents proposed ideas about how to model government-led initiatives after specific projects that were viewed as successful collaborations with academia or industry. Examples included the "strong tradition of government-led super-computing projects, at much larger scale than current AI research" in physics research, as well as "the example J. C. R. Licklider set with his management of the ARPA project." A few offered more general thoughts on how the government should approach AI issues at a broader level.

"It is important that these resources be widely available, not just to those at the top institutions or institutes. Good ideas can come from many places."

But other respondents expressed a lack of confidence in the success of government-built projects, as reflected by one who observed, "I'm very wary (from experience) of using directly U.S. government–supplied resources or staff—I worry that neither will be organized well to support research, and that the reporting overhead will be overwhelming."

In short, these free responses give some insight into the broad diversity of perspectives from AI researchers on government provision of AI research resources. Most who provided responses welcomed more government involvement in AI research, although a number expressed skepticism about the government's capability to provide useful resources. While many did underscore the potential value of compute resources, others emphasized that they viewed workforce development as a clearly higher priority, or suggested that simply scaling up compute resources would not reliably generate new breakthroughs. And among commentators who gave specific recommendations on the creation of government resources, the most common theme was to underscore the importance of making resources accessible and equitable.

Specific research groups may vary in compute needs

To better understand how different subsets of AI researchers view the role of compute in their work, we separated respondents into several groups for further analysis. To supplement our earlier analysis, we contrasted "high" and "low" compute users, as based on reported GPU hours used in a respondent's most compute-intensive project. We also examined the views of language modelers, researchers at AI startups, and academics who rely exclusively on cloud-based computing resources. For some of these groups, we had a small number of responses, so this is a preliminary look at whether the views of these groups diverge from the views of other researchers in our sample. The goal in performing this analysis was to identify populations that may have special compute needs, so that further research can more effectively study how various populations would benefit differentially from increases in access to computing resources.

Defining high compute users as those who report more than 5,000 GPU hours (n = 94), we find that these researchers are more inclined to want compute resources from the federal government. In fact, 67 respondents in this group want government-provided

compute, compared to only 38 low compute users (who report 50 or fewer GPU hours, n = 74).* More high compute users work in computer vision (CV) or natural language processing (NLP), while low compute users more often work in other subfields, such as recommender systems or algorithmic and architectural analysis.[34]

Comparing preferences of high and low compute users, fewer (n = 8) low compute users would prioritize additional compute if afforded a larger budget, compared to high compute users (n = 24). Low compute users are less concerned about compute constraining their ability to make future contributions to the field, with 27 respondents of 74 not at all concerned, compared to only 13 out of 94 high compute users who are not at all concerned. These results reflect the earlier finding that researchers who use more compute tend to be more concerned about future access to compute. This again suggests that making more compute available to all researchers could actually stratify existing differences in use if only high compute users make use of it.

Beyond high and low compute users, we looked at the views of language modelers. Language modeling is a subfield within NLP that is often cited in the conversation around compute usage and AI development, both because it is very attention-grabbing—GPT-4 and ChatGPT are both language models—and because it can be notoriously compute-intensive. Just under half of our respondents who reported working in NLP (n = 143) also reported working on language modeling (n = 70). In both industry and academia, roughly half of NLP researchers reported working on language modeling.

Language modelers and other NLP researchers reported similar compute use, except on the extremes, where more language modelers reported the highest compute usage and no language modelers report the lowest compute usage. There are 10 language modelers in the two highest use categories (GPU hours > 50,000), as compared to only 4 non–language modeler NLP researchers, so these researchers represent a tiny fraction of all those surveyed. While our sample size for this specific population was too small to make general claims about language modelers' views on compute, other survey research has identified interesting divisions among NLP researchers, and further research could address this more explicitly.[35]

---

* High compute users include the top three response categories: 5,001–50,000 GPU hours, 50,000–500,000 GPU hours, and more than 500,000 GPU hours. Low compute users include the lowest two response categories, no GPU hours, and 0–50 GPU hours. We get similar results if cost in dollars is used to define high and low compute users rather than GPU hours. Other desired resources are selected at similar rates between high and low compute users.

Another AI researcher profile of interest is those in industry working at less established—and perhaps less well-resourced—AI startups, which we define as those respondents in our survey who indicated working for an organization in industry with fewer than 500 employees (n = 35). This group includes primarily CV and NLP researchers. While talent was still the top budget priority for this group (n = 15), startup researchers seemed uniquely interested in more data resources (n = 14). By comparison, among all respondents, talent was the top budget priority (n = 223), but data was a much lower priority (n = 93). Additionally, this group more often attributed project success to data, and reported less concern about compute impacting their future contributions to AI. This suggests that data may be a relatively greater obstacle for researchers at AI startups, compared to other types of AI researchers; however, it is unclear if this would remain the case with more respondents from this category.

A final profile we explored was that of academics who rely exclusively on cloud computing for their research (n = 40). We speculated that views on compute may differ based on the type of compute resource on which researchers rely, and that academics who already heavily use cloud computing resources would more immediately benefit from national AI research resources, which are likely to provide compute access via a national cloud resource. This group does appear to be more concerned about compute: a large fraction (n = 16) of these respondents cited compute as a top budget priority when compared to the number of total academics (n = 64 out of 275 academics), while citing talent less often. There is also some indication this group was more concerned about how compute will impact their ability to contribute to the field in the future, in that more respondents were extremely concerned compared to all academic respondents.

This suggests, promisingly, that the researchers who may be most used to using cloud-based resources (and therefore who may benefit most immediately from national resources) are also more likely to want greater compute resources. However, we also observe that this population contains almost no robotics researchers: only 2 out of 40 researchers in this demographic work in robotics as compared to 58 out of 275 academics overall. This discrepancy may be because robotics research requires on-premise compute in a way that other fields do not, which underscores that cloud-like resources will not necessarily benefit all fields of research equally.

Conclusions

Policies designed to promote U.S. AI progress and competitiveness increasingly focus on compute as a primary lever of influence.[36] Current proposals for the NAIRR focus on compute as a primary resource that the government can provide to spur innovation in AI research and increase the diversity of researchers contributing to AI progress.[37] This focus on compute is generally justified by a number of factors, including the following:

1. **Relevance to fundamental AI progress:** While data, compute, algorithms, and talent are all important in machine learning, commentators often note that many algorithms underpinning today's most advanced AI models are decades old. By contrast, since 2012 the amount of compute used by major "notable" AI models has grown shockingly quickly.[38] Some researchers increasingly frame compute as the most relevant constraint facing AI engineers, who may plan their dataset utilization around their compute budget.[39]

2. **Inequity in the status quo:** Proposals for the NAIRR or a NAIRR-like resource that heavily focuses on compute are also frequently justified in terms of the need to mitigate "compute divides" between well-resourced researchers and poorly resourced ones.[40] Consistent with this view, a stated goal of the NAIRR is "to democratize access to research tools that will promote AI innovation and fuel economic prosperity."[41]

3. **Ease of provisioning:** Independent of the value of compute as a contributor to AI progress relative to other input factors, it is reasonable to think that compute resources may be the easiest for the federal government to provide to researchers. Workforce development initiatives may take decades to mature. Immigration reforms to permit more high-skilled immigrants to work in AI development will require congressional approval. Government-provided data may not be appropriate for many important areas of AI research, and its curation can also be resource-intensive while raising legal and privacy concerns.[42] By contrast, acquiring compute is relatively straightforward for the government to do, as well as being relatively straightforward for researchers to use.

The results of this survey cast doubt on the first and second of these justifications, without undermining the third.

With respect to the claim that computing power is the most relevant resource for AI progress, respondents in our sample appeared to disagree. More researchers reported strongly agreeing that compute was a major driver of the past decade of AI progress than were other factors. But larger proportions viewed most other factors we asked

about as likely to be an important driver of the next decade of AI progress as compared to compute. In addition, when regarding their own experiences, respondents reported adjusting their research plans due to a lack of data or talent more often than a lack of compute. And given more funding, most researchers would choose to spend it on talent, not compute. Certain types of AI research are absolutely constrained by compute. Most notably, large "foundation models" tend to be highly compute-intensive, and progress toward larger such models is constrained by compute at present.[43] But based on our sample, the results suggest that these issues affect a small minority of AI researchers.

With respect to computing resource divides, we do find weak evidence that industry and academia have differential access to compute resources, but find stronger evidence that reported compute use is similar across these groups. We also find that a lack of compute resources is not a primary concern in motivating academics to consider leaving for industry, and that academics do not on average report greater concern about their future compute access than do industry researchers.

While our results suggest less of a dramatic difference between the capabilities of industry and academic researchers than is often assumed to exist, there is nonetheless plenty of variation in compute use among respondents in our sample. Motivated by the assumption that most AI researchers want to work on compute-intensive projects, policymakers have at times assumed that such variation is the result of different levels of access to compute.[44] This leads naturally to the conclusion that provisioning large amounts of compute to a wider number of researchers could "democratize" AI research by allowing poorly-resourced researchers to compete with better-resourced ones.

Such a strategy could actually backfire, resulting in differences in compute usage becoming even further stratified. Across a wide number of indicators, we found that the researchers who were most eager for greater amounts of compute were the same ones who already used more compute than their peers. This finding suggests that, rather than being a result of barriers to access, variation in existing compute usage may better be explained by self-selection effects.[*] And this in turn suggests that if more compute were made available across the board to researchers, it might primarily benefit high compute users, without becoming a major resource for researchers currently using less compute. If policymakers view compute-heavy research as more important to promote than less compute-heavy methods, or if they are concerned

---

[*] These self-selection effects could operate at multiple levels: researchers self-select into specific fields, for instance, but they also self-select into using different methodological approaches within those fields. In AI, both fields and the common research methods used within them can vary widely in terms of compute requirements.

about democratizing access to—as opposed to actual use of—compute, then these concerns need not affect proposed policy solutions. But our results suggest that it is uncertain that such policies would address the "lack of diversity" among AI researchers,[45] as opposed to simply further entrenching researchers and methods that currently dominate the field of AI.

Despite these results, it remains true that compute may be relatively easier to provide to AI researchers by comparison with data or talent support, in both legal and logistic terms. Our respondents did indicate that when it comes to government resources, they would be more receptive to compute than government-curated data or technical staff—though they would generally prefer grant funding to compute resources. These considerations suggest that compute may still be an appropriate focus for federal policymaking, whether the goal is to provision resources to researchers or to find appropriate levers for constraining adversary innovation.

At the same time, even if compute is still relatively useful for these aims, policymakers would do well to manage their expectations regarding the overall impact of AI policies that target compute. A focus on compute among policymakers is not riskless. On the domestic side, government-provisioned compute resources could risk further centralizing the economic power of a few small cloud providers or hardware manufacturers.[46] At sufficient scales, these resources would also significantly increase the carbon emissions of the AI industry at a time when such emissions are increasingly a source of concern.[47] Meanwhile, from an international competitiveness angle, policies that heavily restrict another nation's access to compute may end up undermining the U.S. semiconductor industry, just as past attempts at export controls in the satellite industry have inadvertently harmed U.S. companies.[48]
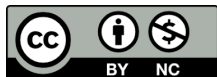
These risks may all be worth taking if access to compute is a primary barrier to breakthroughs in AI, and if increases in compute availability reliably lead to AI dominance. The results of this survey do not disprove this possibility. But they do suggest that such views may not be as widely shared among AI researchers as policymakers often assume.

Authors

Micah Musser is a research analyst at CSET, where he works on the CyberAI Project. Rebecca Gelles, Ronnie Kinoshita, and Catherine Aiken all work for CSET's Data Team, respectively as a data scientist, a survey research analyst, and the director of data science and research. Andrew Lohn also works on the CyberAI Project at CSET as a senior fellow.

Acknowledgments

Appendices

A. Sampling Methodology

Top AI Conferences and Journals

We included a conference or journal in this list if it met either of the following two criteria as of (approximately) March 1, 2022:

1) It was tracked by the website CSRankings.org as a top conference or journal in the subfields of artificial intelligence, computer vision, machine learning and data mining, or natural language processing;

2) The conference or journal had an h5-index over 100 as tracked by Google Scholar in the subfields of artificial intelligence, computational linguistics, computer vision and pattern recognition, data mining and analysis, or robotics.[49]

Table A.1 shows the final list of included conferences and journals, along with the number of results from each one.

Table A.1. Number of Authors Identified in AI Conferences and Journals

Conference or Journal Name (Abbreviation on CSRankings, if tracked)	Number of Results
AAAI Conference on Artificial Intelligence (AAAI)	5,953
International Joint Conference on Artificial Intelligence (IJCAI)	1,178
IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	5,010
European Conference on Computer Vision (ECCV)	1,189
IEEE International Conference on Computer Vision (ICCV)	2,431
International Conference on Machine Learning (ICML)	2,686
International Conference on Knowledge Discovery and Data Mining (KDD)	2,266
Neural Information Processing Systems (NeurIPS/NIPS)	4,547
Annual Meeting of the Association for Computational Linguistics (ACL)	2,585

Conference on Empirical Methods in Natural Language Processing (EMNLP)	886
North American Chapter of the Association for Computational Linguistics (NAACL)	72
International Conference on Learning Representations	91
IEEE Transactions on Systems, Man, and Cybernetics Part B—Cybernetics	0
Expert Systems with Applications	900
IEEE Transactions on Neural Networks and Learning Systems	781
Neurocomputing	548
Applied Soft Computing	242
IEEE Transactions on Pattern Analysis and Machine Intelligence	1,159
IEEE Transactions on Image Processing	881
IEEE International Conference on Robotics and Automation	4,916

Source: CSET

AI Roles in LinkedIn Profiles

We identified AI researchers working in industry if their LinkedIn profile met either of the following criteria as of (approximately) March 1, 2022:

1) The respondent's current role on LinkedIn was listed as machine learning engineer, machine learning architect, machine learning analyst, machine learning lead, artificial intelligence engineer, artificial intelligence architect, artificial intelligence analyst, or artificial intelligence lead;

2) The respondent's current employer on LinkedIn was listed as an AI startup included on the CB Insights list of 46 AI startups, and the respondent's current role was listed as one of the job titles in Table A.2, which follows.[50]

Table A.2. List of Job Titles Used to Identify AI-Relevant Employees on LinkedIn

Job title		
Advisory Software Engineer	Information Analyst	Scientist
Analyst Programmer	Infrastructure Analyst	SDE
Analytics Specialist	Infrastructure Architect	Software Designer
Automation Engineer	Infrastructure Engineer	Software Developer
Cloud Architect	Java Developer	Software Engineer
Data Analyst	Machine Learning Engineer	Statistical Programmer
Data Analytics	Programmer Analyst	Statistician
Data Architect	Quantitative Analyst	Technical Architect
Data Center Operator	Research and Development Engineer	Technical Lead
Data Engineer	Research and Development Specialist	Technical Product Manager
Data Scientist	Research and Development Engineer	Technical Project Manager
Development Engineer	Research and Development Specialist	Technology Lead
	ETL Developer	Researcher

Source: CSET

In addition to the preceding selection criteria, our respondents were screened at the beginning of the survey to ensure that they built, developed, studied, or maintained AI systems "at least some of the time."

B. Response Rates and Sample Representativeness

Our final response rate of 1.7 percent reflects other recent web-based surveys of AI researchers and experts, which have reported response rates from 1.1 to 21 percent, as shown in Table B.1. Several of the surveys either offered incentives or had particularly large samples from academia, both of which can increase response rates.

Table B.1. Response Rates from Recent Surveys of AI Researchers and Experts

Author	Population	Distribution Method	Incentives Offered?	Response Rate
OECD (2023)[51]	"An audience with expertise or knowledge of AI compute"	Online survey; precise distribution unclear	No	N/A
Michael et al. (2022)[52]	Active members of the Association of Computational Linguistics	ACL membership mailing list; in-person ACL events; Twitter; Slack; email distribution	Yes	5%
RAND (2022)[53]	Software engineers (Silicon Valley employees/alumni of top CS universities)	Email distribution; LinkedIn advertisements; Northrop Grumman AI Academy	No	1.1%
Zhang et al. (2021)[54]	AI researchers with at least two prominent publications	Email distribution	Yes	17%
CSET (2020)[55]	AI PhD graduates with AI-relevant dissertations from top-ranking universities	Email distribution	No	11%
Grace et al. (2018)[56]	Researchers who published at the 2015 NeurIPS and ICML conferences	Email distribution	Yes	21.5%

Source: CSET

As with most surveys, not all who were invited chose to participate. This raises concerns regarding nonresponse bias, which can occur if respondents differ meaningfully from those who choose not to respond on characteristics relevant to the study. If present, nonresponse bias threatens the validity of conclusions drawn from the survey. It is worth noting, however, that a low response rate does not itself introduce nonresponse bias; response rates can be low without meaningful differences between respondents and nonrespondents, while response rates can be high and have meaningful differences between those groups.[57]

One way of checking for nonresponse bias is to compare respondents and nonrespondents along known characteristics, which in our survey included each respondent's sector (academia, industry, or government) as indicated by the domain of their email address or by their self-identification within the survey. Table B.2 summarizes the total number and sector breakdown of respondents who received, began, and completed the survey. While the final column is based on self-identification within the survey, the first and second columns are based on email domains. Academics are defined as respondents with an ".edu" email address, government respondents as those with a ".mil" or ".gov" email address, and industry respondents as the remainder.

Table B.2. Sector of Respondents Who Received, Began, and Completed the Survey

Sector	Composition among researchers who:		
	received an email with the survey link[*]	began the survey[†]	completed the survey[‡]
Industry	38% (11,575)	31% (195)	29% (120)
Government	2% (493)	3% (16)	3% (14)
Academia	60% (18,243)	67% (423)	67% (275)
Total	30,311	634	410

Source: CSET Compute Resource Survey

With respect to sector composition, we do observe a statistically significant difference ($\chi^2 = 19.96$, $p < 0.001$) between nonrespondents and the sample of researchers who completed the survey. In particular, academics who received our survey were more likely to complete it, while industry respondents were less likely to do so. We were not able to evaluate nonresponse bias in terms of unobservable characteristics, and for other characteristics about which our survey did ask—such as field of study—we were not able to compare the composition of our respondents to the composition of our overall sampling frame. We considered weighting our responses to account for the observed nonresponse bias in terms of sector, but chose not to do so. For Findings 2.1, 2.2, and 2.3, the analysis in question either only included academics or directly compared academics with industry respondents, in which case weighting based on sector would be irrelevant. For Findings 1.1, 1.2, 1.3, and 2.4, academics generally expressed slightly more concern regarding compute; weighting to account for the greater nonresponse rate among industry respondents would therefore make our core results appear even stronger than we present them here.

[*] Percentages calculated as estimates based on email domains. If the email failed, bounced, or was blocked before delivery, it is not included in this column.

[†] Percentages calculated as estimates based on email domains. These counts are based on a Qualtrics distribution report, which excludes seven respondents who only partially completed the survey, and includes respondents who self-screened out by indicating they did not consent to participate in the study.

[‡] Percentages based on respondent's identification within the survey itself. Percentages do not add up to 100% because one respondent listed "None of these" as primary affiliation.

C. Subfield Comparisons

Respondents in our survey were asked to indicate whether they worked in the five top-level fields of computer vision, NLP, robotics, reinforcement learning (RL), and "other." Respondents who indicated working in any of these top-level fields were then shown a series of subfields related to the top-level category and asked to indicate if they worked in each of those subfields.[*] On average, respondents in each of the five top-level categories indicated working in roughly a quarter of the related subfields independently of the top-level category in question, as shown in Table C.1. In addition, the median academic reported working in a total of three subfields (not including the five top-level fields and the final "none of these" option), while the median industry researcher reported working in a total of four subfields.[†] Figure C.1. shows the number of subfields indicated by industry and academic respondents for each top-level field.

Table C.1. Number and Percent of Subfield Options Selected by Respondents in Each Field

Field	Mean Number of Subfields Selected	Number of Subfields Presented	Mean Percent of Subfields Selected
Computer Vision	2.82	10	28%
Robotics	2.74	9	30%
Natural Language Processing	4.22	17	25%
Reinforcement Learning	1.84	8	23%
Other	1.78	7	25%

Source: CSET Compute Resource Survey

---

[*] A full list of the subfields can be viewed in either the survey instrument or the file "data/field_composition.csv" within this project's GitHub repository, accessible at https://github.com/georgetown-cset/Compute_Survey_2022.

[†] The difference between the median number of subfields indicated by academics as opposed to industry researchers is significant at $p = 0.004$ according to a Mann-Whitney U test. However, this difference is likely explained by the fact that more respondents from industry than from academia reported working in NLP (see the footnote on page 10, above), and substantially more subfields were presented to respondents in NLP than in other fields.

Figure C.1. Number of Subfields Selected by Researchers by Sector
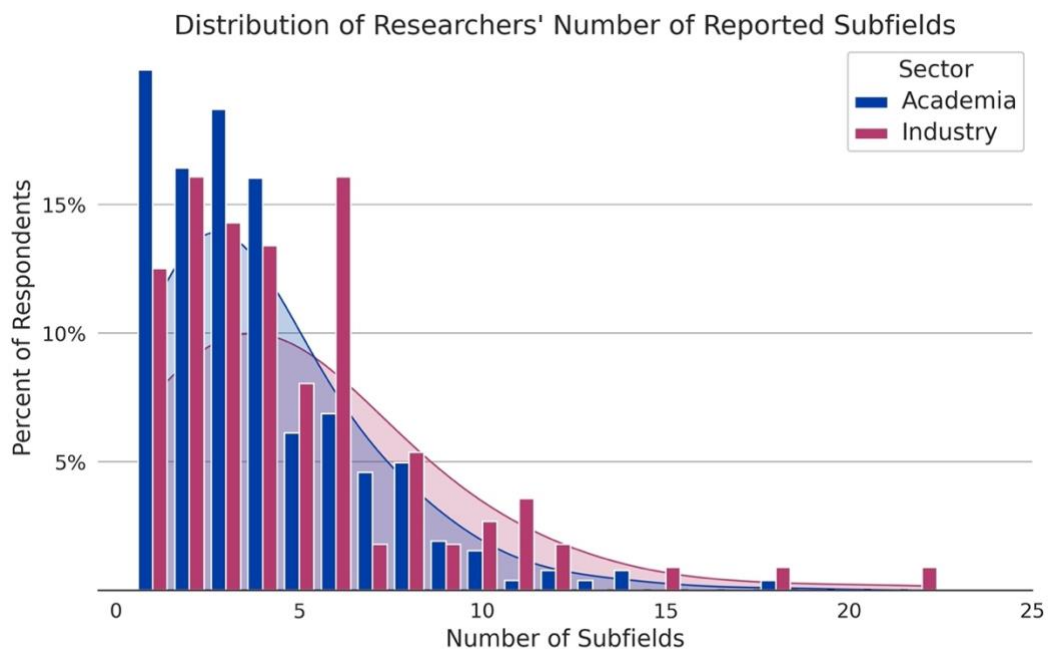
Source: CSET Compute Resource Survey

Figure C.2 provides some additional insight into the variation among these subfields. Each point in this figure represents one subfield, with the location on the x-axis indicating the average compute expenditure on a researcher's most compute-intensive project reported across all researchers in the subfield, and the location on the y-axis indicating the mean level of concern from researchers that future contributions will be limited by a lack of access to compute. The positive correlation between the two variables further reflects Finding 2.4 of this report: that researchers who already use larger quantities of compute tend to report being more concerned about their lack of access to compute. However, Figure C.2 also illustrates that researchers across subfields in both computer vision and NLP are fairly tightly clustered together in a high-compute-use and high-concern category,[*] while the subfields of robotics and reinforcement learning exhibit much higher variance. Robotics in particular exhibits significantly less concern that future research will be limited by compute, relative to other subfields.

---

[*] The tight clustering of subfields within NLP may be due to the fact that respondents working in NLP indicated, on average, working in a larger number of subfields than in other top-level fields; see Table C.1.

Figure C.2. Mean Compute Use and Concern over Future Compute across Subfields
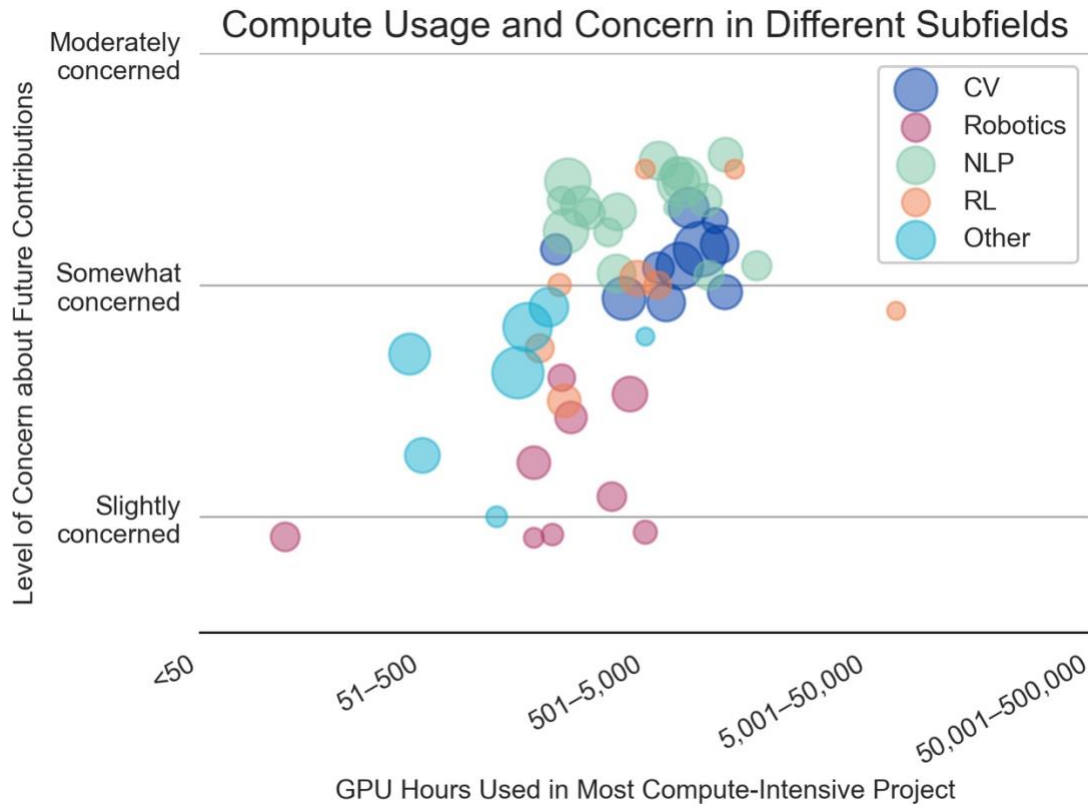
Source: CSET Compute Resource Survey

Finally, Figure C.3 shows that the percent of respondents indicating a desire for government-provisioned compute resources varies substantially by subfield (dotted black lines show the percent across all respondents in a top-level field indicating such a desire). In general, NLP and computer vision researchers are the most likely to indicate support for such resources, and robotics researchers are the least likely. However, there is meaningful variation within CV and RL in particular, with not all subfields equally interested in using government-provided compute resources.

Figure C.3. Percent of Respondents Indicating a Desire for Government-Provided Compute by Subfield
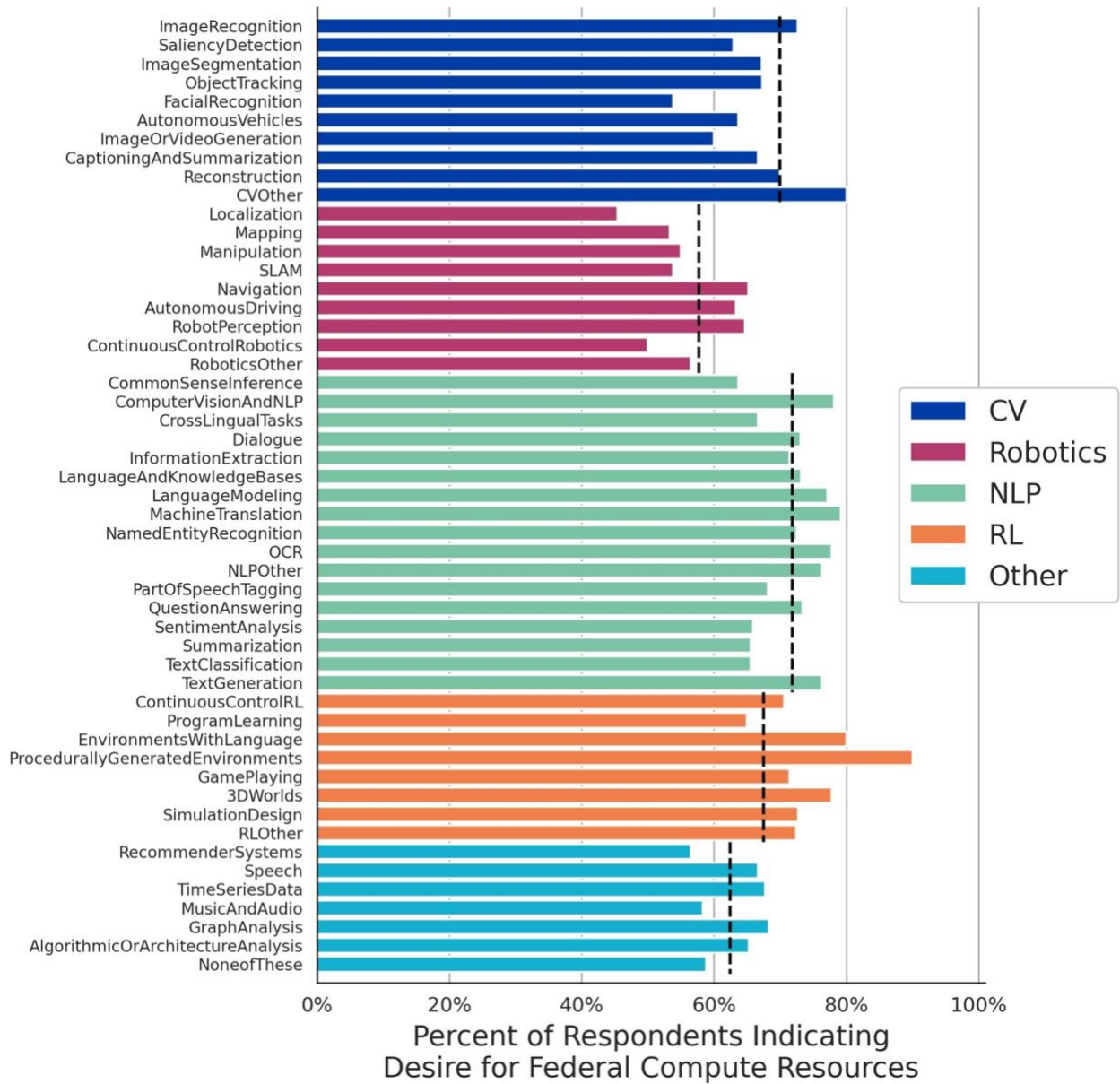
Source: CSET Compute Resource Survey

Endnotes

[1] Eric Schmidt, Robert Work, Safra Catz, Eric Horvitz, Steve Chien, Andrew Jassy, and Mignon Clyburn, et al., "Final Report" (National Security Commission on Artificial Intelligence, October 2021), 185. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

[2] Schmidt et al., "Final Report," 186–87. See also Dario Amodei and Danny Hernandez, "AI and Compute," *OpenAI*, May 16, 2018, https://openai.com/research/ai-and-compute; and Nur Ahmed and Muntasir Wahed, "The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research," arXiv [cs.CY], October 22, 2020, https://doi.org/10.48550/arXiv.2010.15581.

[3] Implementation of Additional Export Controls: Certain Advanced Computing and Semiconductor Manufacturing Items; Supercomputer and Semiconductor End Use; Entity List Modification, 87 Fed. Reg. 62,186 - 62,215 (Oct. 7, 2022) (revising 15 C.F.R. §§ 734–774).

[4] Bureau of Industry and Security Office of Congressional and Public Affairs, "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)," press release, October 7, 2022, https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file; Jake Sullivan, "Remarks by National Security Advisor Jake Sullivan at the Special Competitive Studies Project Global Emerging Technologies Summit," The White House, September 16, 2022, https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/09/16/remarks-by-national-security-advisor-jake-sullivan-at-the-special-competitive-studies-project-global-emerging-technologies-summit/.

[5] "These controls were imposed because GPU chips play an important role in the development and use of artificial intelligence applications, particularly the deep learning methods that are the main driver of the current AI boom." Martijn Rasser and Kevin Wolf, "The Right Time for Chip Export Controls," *Lawfare*, December 13, 2022, https://www.lawfareblog.com/right-time-chip-export-controls. "By only targeting chips with very high interconnect speeds, the White House is attempting to limit the controls to chips that are designed to be networked together in the data centers or supercomputing facilities that train and run large AI models." Gregory C. Allen, "Choking off China's Access to the Future of AI" (Center for Strategic & International Studies, October 11, 2022), https://www.csis.org/analysis/choking-chinas-access-future-ai.

[6] Office of Science and Technology Policy, "The Biden Administration Launches the National Artificial Intelligence Research Resource Task Force," press release, June 10, 2021, https://www.whitehouse.gov/ostp/news-updates/2021/06/10/the-biden-administration-launches-the-national-artificial-intelligence-research-resource-task-force/.

[7] National Artificial Intelligence Research Resource Task Force (NAIRR), "Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource," January 2023, https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf, 48–49.

[8] NAIRR Task Force, "Strengthening and Democratizing," 7. See also the discussion of "compute divides" in, among other sources, "A Blueprint for Building National Compute Capacity for Artificial Intelligence," *OECD Digital Economy Papers* no. 350 (February 2023), https://oecd.ai/en/compute-report; and Daniel E. Ho, Jennifer King, Russell C. Wald, and Christopher Wan, "Building a National AI Research Resource: A Blueprint for the National Research Cloud" (Stanford Institute for Human-Centered Artificial Intelligence, October 2021), 19–20, https://hai.stanford.edu/sites/default/files/2022-01/HAI_NRCR_v17.pdf. The first of these documents explicitly argues that the current degree of attention paid to compute resources is insufficient: "While other key enablers have received significant attention in policy circles, the hardware, software, and related compute infrastructure that make AI advances possible receive comparatively less attention." "A Blueprint for Building National Compute Capacity," *OECD Digital Economy Papers*, 13. While this strategy document argues that current "compute divides" are deeply worrying, it also admits that current measurement frameworks make it impossible to evaluate how serious such divides are.

[9] See Amodei and Hernandez, "AI and Compute"; Andrew Lohn and Micah Musser, "AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?" (Center for Security and Emerging Technology, January 2022), https://doi.org/10.51593/2021CA009; and Jaime Sevilla, Lennart Heim, Anson Ho, Tamay Besiroglu, Marius Hobbhahn, and Pablo Villalobos, "Compute Trends across Three Eras of Machine Learning," arXiv [cs.LG], March 9, 2022, https://doi.org/10.48550/arXiv.2202.05924.

[10] Ahmed and Wahed, "The De-democratization of AI."

[11] See Amodei and Hernandez, "AI and Compute"; Lohn and Musser, "AI and Compute"; and Sevilla et al., "Compute Trends across Three Eras."

[12] Web of Science is provided by Clarivate Analytics, https://clarivate.com/webofsciencegroup/solutions/web-of-science/.

[13] RocketReach, https://rocketreach.co/.

[14] Response rate calculated using the American Association for Public Opinion Research, "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 9th edition," 2016, https://aapor.org/standards-and-ethics/standard-definitions/. Three pilot versions of the survey were sent to random samples of 500 identified AI researchers in late spring 2022. These pilot surveys were used to estimate a likely response rate and to further improve the survey instrument. Primary survey distribution occurred in June 2022. A final follow-up distribution was conducted in July 2022 to 50 AI researchers who had not received the survey in previous distributions due to invalid emails. We manually identified alternate emails for those individuals and sent them the survey. Responses from the pilot and follow-up distributions are included in the analysis. Partial responses are included in the analysis only where the relevant question was completed; we do not report rates of nonresponse in percentage breakdowns.

[15] Our five cognitive interviewees were carefully selected to represent a range of backgrounds, including a tenured professor, two PhD students, and technical engineers from a major tech company and an AI lab. The interviewees came from multiple fields of AI research, including natural language processing,

data mining, and robotics. During cognitive interviews, we evaluated how interviewees interpreted and responded to different questions on the survey, and we examined response time and cognitive burden of individual questions and the full survey. These interviews informed minor revisions to the survey instrument.

[16] Among respondents indicating that their most compute-intensive project was their most significant project, 62 percent indicated that compute was very or extremely important to the project's success. It is possible that a researcher's most compute-intensive project was also their largest project in other ways—in terms of time, cost, data requirements, or personnel—but we did not ask whether the project in question was also the researcher's most intensive project on any of these other dimensions.

[17] All pairwise comparisons were made using Mann-Whitney U tests with 3-way Bonferroni correction. Differences between compute and data were significant for rejecting, revising, and abandoning projects (all p < 0.001). Differences between compute and researcher availability were significant for rejecting and abandoning projects (both p < 0.001), but not for revising projects (p = 0.773). Differences between data and researcher availability were significant for revising projects (p = 0.038), but not for rejecting (p = 1.0) or abandoning projects (p = 0.497).

[18] "A Blueprint for Building National Compute Capacity," *OECD Digital Economy Papers*, 60.

[19] "Other AI enablers, like data, algorithms, and skills, receive significant attention in policy circles, but the hardware, software, and related infrastructure that make AI advances possible have received comparatively less attention." "A Blueprint for Building National Compute Capacity," *OECD Digital Economy Papers*, 5.

[20] Academic researchers can oftentimes replicate results from industry, even when models cost very large sums to train initially. For instance, after the announcement of AlphaFold 2, but prior to its release by DeepMind as an open-source model, a team of researchers at the University of Washington developed an alternative open-source model, RoseTTaFold, which performed nearly as well as AlphaFold. Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, and Jue Wang, et al., "Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network," *Science* 373, no. 6557 (Aug 2021): doi.org/10.1126/science.abj8754.

[21] Richard Sutton, "The Bitter Lesson," accessed October 6, 2022, http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

[22] Lohn and Musser, "AI and Compute."

[23] Difference in distribution for reported compute use in dollars between the two groups is statistically significant, with p < 0.001 by a Mann-Whitney U test. The difference between the two groups for reported compute use in GPU hours is not statistically significant by a Mann-Whitney U test (p = 0.225). This project's GitHub repository (https://github.com/georgetown-cset/Compute_Survey_2022/) contains the results of two ordinal logistic regression models that further analyze these results. The first model (Model 1) suggests that computer vision researchers and NLP researchers are more likely to report higher GPU usage for their most compute-intensive project than are other types of researchers (p < 0.001 and p = 0.071, respectively) after adjusting for differences across sectors. There are significantly

more NLP researchers in industry in this sample than in academia (see the footnote on page 10, above), so differences between fields could explain the slight difference between industry and academia for GPU hours. However, Model 2 shows that even after accounting for differences in field, industry researchers report significantly higher compute usage in monetary terms ($p < 0.001$).

[24] For instance: if academics report spending less on compute than industry researchers but do not report using fewer GPU hours on average, this may signal that academics are purchasing access to cheaper (and lower-performing) GPUs than those used in industry.

[25] It could be argued that dollars are a more salient metric for researchers, or for researchers in academia specifically, and thus the more valid metric. That more respondents were able to report GPU hours suggests that may not be the case or may be changing. Note that 18 percent of industry researchers and 9 percent of academics did not report compute use by either metric.

[26] Among academics, 82 percent reported using on-premise compute (46 percent reported using exclusively on-premise compute), and 52 percent of industry researchers reported using on-premise compute (22 percent exclusively). The differences in the proportions of both on-premise compute users and exclusive on-premise compute users between academic and industry respondents were statistically significant by a chi-squared test of independence at $p < 0.001$.

[27] Roman Jurowetzki, Daniel Hain, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos, "The Privatization of AI Research(-ers): Causes and Potential Consequences—From University-Industry Interaction to Public Research Brain-Drain?" arXiv [cs.CY], February 2, 2021, https://doi.org/10.48550/arXiv.2102.01648. Note that we did not ask researchers in private industry whether they had seriously considered leaving industry to work in academia. As well, other CSET research cautions against an overemphasis on the "brain drain" narrative as a description of current trends in academic hiring for AI positions. See Remco Zwetsloot and Jack Corrigan, "AI Faculty Shortages: Are U.S. Universities Meeting the Growing Demand for AI Skills?" (Center for Security and Emerging Technology, July 2022), https://doi.org/10.51593/20190049.

[28] Catherine Aiken, James Dunham, and Remco Zwetsloot, "Career Preferences of AI Talent" (Center for Security and Emerging Technology, June 2020), https://doi.org/10.51593/20200012. This 2019 survey of graduates from top AI PhD programs found that only 31 percent of respondents rated "access to compute resources or interesting data" as extremely important in choosing where to work, by comparison with 46 percent for "financial compensation" and 64 percent for "interesting technical challenges." This survey only asked respondents whether each factor was "not at all important," "somewhat important," or "extremely important," as opposed to using the five-item scale featured in the present research.

[29] To determine this, we constructed a contingency table representing the number of respondents in both academia and industry who either (1) reported greater growth in compute needs than compute access (column 1 of Figure 8); or (2) reported greater growth in compute access than compute needs, or reported the same value for both (columns 2 and 3 of Figure 8). By a chi-squared test of independence, academics were significantly more likely to report growth in compute needs that was not matched by growth in compute access ($p = 0.004$).

30 A Mann-Whitney U test results in a p-value of 0.123. We do find that there were significant differences in the field makeup of respondents from academia and those from industry (see the footnote on page 10, above); comparing responses of academics and industry researchers will not show differences in level of concern that may exist at the field level. Model 3 in this project's GitHub repository addresses this with an ordinal logistic regression of level of concern on sector, research field, and the interaction between sector and research field. This model does not result in significant differences between academics' level of concern about their access to compute and that of industry researchers, even when taking field into account.

31 Models 4 and 5 in this project's GitHub repository use ordinal logistic regression to regress respondents' level of concern regarding future compute access against sector, GPU usage for the respondent's most compute-intensive project, and the interaction between these two variables (where Model 4 treats GPU utilization as a categorical variable and Model 5 treats it as a linear variable). While both models identify a significant positive correlation between GPU usage and stated concern about future contributions, neither detects a significant impact of sector, either alone or in interaction with GPU usage.

32 However, note that "technical staff" is a somewhat narrower category than talent more broadly, and that other questions used broader definitions of talent when trying to evaluate how much researchers felt constrained by talent relative to other factors such as compute. We used the narrower category in the context of this question because a national AI research resource would not be positioned to directly hire researchers on behalf of resource users.

33 To analyze these responses, we engaged in focused coding and comparison methods. We documented ideas, questions, and comments, and created a list of themes. A codebook was developed on the basis of agreement between two members of the research team. After initial familiarization with the responses, two members of the research team independently coded responses using thematic analysis. Themes were then collapsed and collated to build out a codebook with a master list of themes. We do not report agreement metrics, as both researchers coded all responses and resolved any discrepancies.

34 When we distinguish high and low compute users by cost in dollars, as opposed to GPU usage, we see a more equal proportion of NLP respondents in both groups, with roughly a quarter of both the highest- and lowest-spending respondents reporting working in NLP.

35 Julian Michael, Ari Holtzman, Alicia Parrish, Aaron Mueller, Alex Wang, Angelica Chen, and Divyam Madaan, et al., "What Do NLP Researchers Believe? Results of the NLP Community Metasurvey," arXiv [cs.CL], August 26, 2022, 10.48550/arXiv.2208.12852.

36 Implementation of Additional Export Controls, 87 Fed. Reg. 62,186 - 62,215.

37 NAIRR Task Force, "Strengthening and Democratizing," 48.

38 Amodei and Hernandez, "AI and Compute"; Sevilla et al., "Compute Trends across Three Eras."

[39] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei, "Scaling Laws for Neural Language Models," arXiv [cs.LG], January 23, 2020, https://doi.org/10.48550/arXiv.2001.08361.

[40] NAIRR Task Force, "Strengthening and Democratizing"; "A Blueprint for Building National Compute Capacity," *OECD Digital Economy Papers*; and Ho et al., "Building a National AI Research Resource."

[41] Office of Science and Technology Policy, "The Biden Administration Launches the National Artificial Intelligence Research Resource Task Force," press release, June 10, 2021, https://www.whitehouse.gov/ostp/news-updates/2021/06/10/the-biden-administration-launches-the-national-artificial-intelligence-research-resource-task-force/.

[42] Ho et al., "Building a National AI Research Resource," 53–60; Amba Kak, Brittany Smith, Sarah Myers West, and Meredith Whittaker, "Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource" (AI Now Institute and Data & Society Research Institute, October 2021), https://ainowinstitute.org/AINow-DS-NAIRR-comment.pdf.

[43] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al., "On the Opportunities and Risks of Foundation Models," arXiv [cs.LG], July 12, 2022, https://doi.org/10.48550/arXiv.2108.07258.

[44] For instance, in calculating expected resource needs for the NAIRR, the NAIRR Task Force "assumed that all federally funded AI researchers throughout the United States from the targeted user communities would use the NAIRR to some extent," and further that "the average computing [*sic*] used by a NAIRR user would be comparable to that of a typical researcher using advanced computing resources." NAIRR Task Force, "Strengthening and Democratizing," 48. This set of assumptions reflects the implicit belief that what prevents individuals and organizations in the "targeted user communities"— which the report defines broadly as including academics, non-profits, government agencies, startups, and small businesses—from using levels of compute in line with "a typical researcher using advanced computing resources" is a lack of access, not a lack of interest in using such compute.

[45] See NAIRR Task Force, "Strengthening and Democratizing," 1.

[46] Ho et al., "Building a National AI Research Resource," 31; Kak et al., "Request for Information."

[47] Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP," arXiv [cs.CL], June 5, 2019, https://arxiv.org/abs/1906.02243; Payal Dhar, "The Carbon Impact of Artificial Intelligence," *Nature Machine Intelligence* 2 (2020): 423–25, https://www.nature.com/articles/s42256-020-0219-9. See also NAIRR Task Force, "Strengthening and Democratizing," 28–29.

[48] Tim Hwang and Emily S. Weinstein, "Decoupling in Strategic Technologies: From Satellites to Artificial Intelligence" (Center for Security and Emerging Technology, July 2022), https://doi.org/10.51593/20200085.

[49] "Top Publications," Google Scholar, accessed March 1, 2022, https://scholar.google.es/citations?view_op=top_venues&hl=en&vq=eng.

[50] "The United States of Artificial Intelligence Startups," *CB Insights*, August 4, 2021, https://www.cbinsights.com/research/artificial-intelligence-startup-us-map/.

[51] "A Blueprint for Building," *OECD Working Papers*. Response rate cannot be calculated due to an unclear sampling frame; it appears that primary distribution of the survey may have been via tweet. In all, 118 responses (both partial and complete) were recorded.

[52] Michael et al., "What Do NLP Researchers Believe?" Response rate is defined as compared to the total population of "active" ACL members, who may or may not have been directly reached by one of the associated distribution methods.

[53] James Ryseff, Eric Landree, Noah Johnson, Bonnie Ghosh-Dastidar, Max Izenberg, Sydne J. Newberry, Christopher Ferris, and Melissa A. Bradley, "Exploring the Civil-Military Divide over Artificial Intelligence" (RAND Corporation, 2022), https://www.rand.org/pubs/research_reports/RRA1498-1.html. Median response time was 15 minutes.

[54] Baobao Zhang, Markus Anderljung, Lauren Kahn, Noemi Dreksler, Michael C. Horowitz, and Allan Dafoe, "Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers," arXiv [cs.CY], May 5, 2021, 10.48550/arXiv.2105.02117. Median response time was 17.4 minutes.

[55] Aiken, Dunham, and Zwetsloot, "Career Preferences of AI Talent." Median response time was 18 minutes.

[56] Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans, "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts," *The Journal of Artificial Intelligence Research* 62 (2018): 729–54. https://doi.org/10.1613/jair.1.11222. Median response time was 12 minutes.

[57] Robert M. Groves and Emilia Peytcheva, "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis," *Public Opinion Quarterly* 72, no. 2 (2008): 167–89. https://doi.org/10.1093/poq/nfn011.