C-ACIS Topic: Computing as Research Infrastructure

Disclaimer: This document can be viewed by any member of the Stanford community (the permissions are set to Stanford only). The purpose of this document is to inform discussions during the 2023-2024 C-ACIS meeting cycle.

This document will be provided to the committee members ahead of time in preparation for discussion at a C-ACIS meeting. Please fill in information for C-ACIS to read before the meeting. This document should include all of the background and information necessary to have a detailed discussion on the topic. You can include links to other documents (e.g., slide decks, other reports) as needed. The C-ACIS coordinators will send you a calendar invitation for the date and time that your topic will be covered.

All of the asterisked* sections are required; the unshaded boxes will expand as you type.

Topic Title*
Provide the title of your topic as it should appear on the C-ACIS agenda. Be brief and descriptive.

Defining Compute Infrastructure as a Research Instrument

Opportunity or Problem Statement*
Articulate the issue that benefits from C-ACIS's discussion as well as why do you want this topic on the C-ACIS agenda? Is it an informational topic? Do you want feedback? etc.

I came across the issue of compute not being defined as a research instrument when submitting a **proposal to the C-Sharp Shared Resource Instrumentation & Enhancement Program**. We proposed a large compute cluster for the new Stanford Robotics Center (customers: 25 faculty across 8 departments. Cost: $2M) that would offer crucial new research and teaching capabilities not supported by existing compute infrastructure and at a scale that would have positioned Stanford for continued success in robotics and machine learning research. The absence of such compute capabilities not only limits our current research, but risks Stanford missing a critical opportunity to pioneer research and development of high-capacity learned models, making us increasingly reliant on the advances of peer institutions and industry partners. **It got instantly rejected because compute is not considered a research instrument** and despite this not being mentioned in the proposal call.

While we wrote this proposal specifically for the Stanford Robotics Center where faculty are mostly affiliated with the School of Engineering, the need for sufficient compute is prevalent across many of Stanford's schools (see supporting materials). Researchers need access to compute to support their world-class research. Educators need access to compute to prepare their students for a successful career. Defining compute infrastructure as a research instrument will ensure that calls for proposals such as the one by C-Sharp will accept compute proposals and consider them fully.

It would be great if C-ACIS can support this measure and provide a recommendation to for example C-RES and the Dean of Research that define programs like C-Sharp.

Solution Statement*
Provide as much information as needed to describe what a good solution or outcome looks like.

A good outcome would be if C-ACIS would recommend defining compute infrastructure as a research instrument. This recommendation could then be brought to C-RES and the Dean of Research to modify programs like C-Sharp accordingly.

Decision or Question*
Provide the specific question or decision for consideration by C-ACIS.

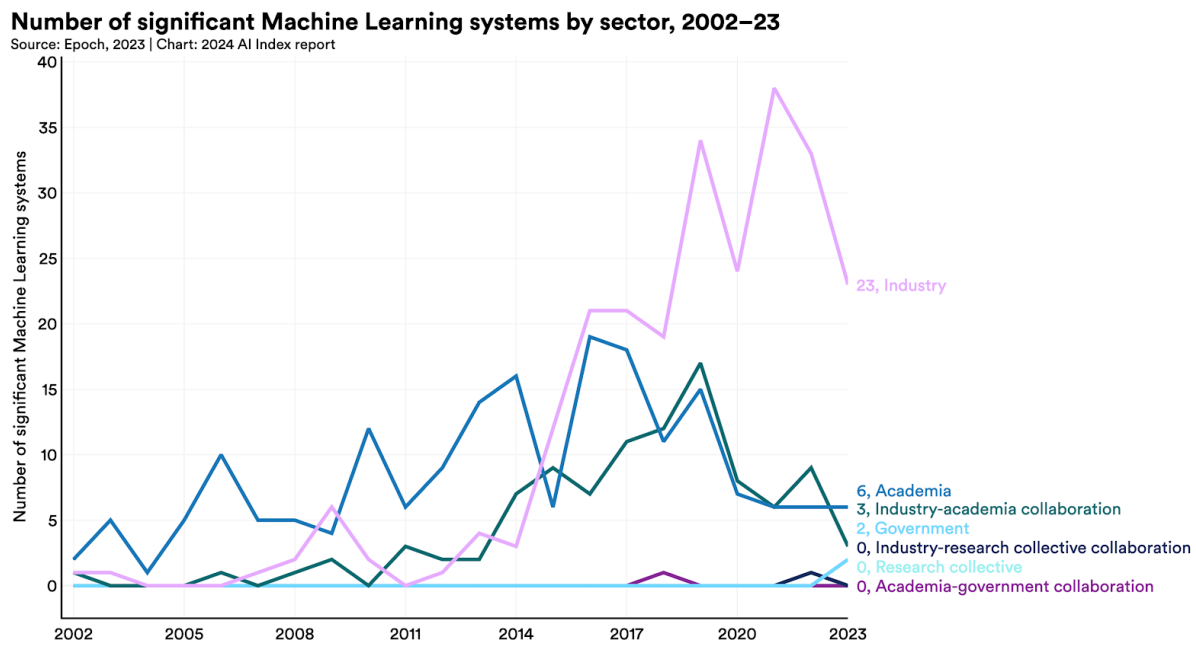
Should compute infrastructure be defined as a research instrument?

Supporting Materials
Please provide any charts, data, reports, or slides that you would like viewed in coordination with this topic. You may insert images directly in this document (i.e., your copy of this template), or you may link to files (please provide access).

Rejected C-Sharp Compute Proposal including proposal, appendix and budget.
HAI slide deck on where their compute grants are going. Key points:

- Slide 3: Since 2020, HAI received about $2m in compute funding which about met the demand across the university. In about summer ’22 demand spiked and HAI can no longer keep up
- Slide 7 shows projects and credit distribution across the university
- Slide 8: is the amount people are requesting across projects and some project examples
- In the call that ended this past Friday, HAI received 49 applications totally $2.1M in requests

Sneak peak from this year's AI Index (see Figure below): Until about 2014 most significant machine learning models were released by academia. Since then, industry has taken over. Of the 6 from academia in 2023, 4 are from Stanford researchers and two were partially funded through the HAI cloud program.



The [c-ShARP Charge] uses this definition:

"Note that we use "shared facilities" and "shared research platforms" to denote any experimental equipment that resides in shared space, spanning SOM, SOE, H&S, and SE3. Example facilities include nanofabrication and characterization (SNF, SNSF), mass spectroscopy (SUMS) and NMR, imaging (c-EMc, Wu Tsai, BioX), life sciences/medicine (ChEM-H), and sustainability, as well as shared Stanford facilities at SLAC."

- Data science is purchasing compute research instrument right now: 300 GPUs for the university -> money from president
- Princeton got 700 GPUs and only for PLI center (not the entire university)
- CTO of Microsoft (last version of GPT was trained on 14000 GPUs)
- HAI is trying to buy 32 GPUs (contracting phase)

Timeframe and Deadlines Provide key dates for the progress of your Solution/Proposal/Initiative/Project.
N/A

Presenters and Attendees Provide names and SUNet IDs of all individuals you would like invited to present and/or attend this topic.	
Names:	**SUNet IDs:**

Jeannette Bohg	**bohg**

Logistics	
Date* Select ALL dates presenters are available.	**Time Expected*** Provide the estimated time needed to discuss. NOTE: We will do our best to accommodate your time request, however we will need to prioritize time based on the number of topics needed through the year.
☐ **October 2023** ☐ **November 2023** ☐ **January 2024** ☐ **February 2024** ☑ ~~**March 2024**~~ ☐ **May 2024**	☐ **15 Minutes** ☑ ~~**20 Minutes**~~ ☐ **30 Minutes** ☐ **45 Minutes** ☐ **60 Minutes** ☐ **N/A - Informational write-up only**