Bootstrapping Data Infrastructure

for Science Policymaking and Communication

Zhane Yamin - AIP Mather Policy Intern with the National Science Board

Zhane Yamin

AIP Mather Public Policy Intern with the National Science Board

What I've been doing this summer, at a glance:

- Data analysis
- Communications
- Strategy
- Networking
- Learning
- A lot of excel

The National Science Board is....

A congressionally established board with a dual mandate to establish the policies of the NSF and to serve as an independent advisor to Congress and the President.

The NSF is the main federal funder of U.S. science & engineering research and talent

And, lots of food...

Year: Rising senior

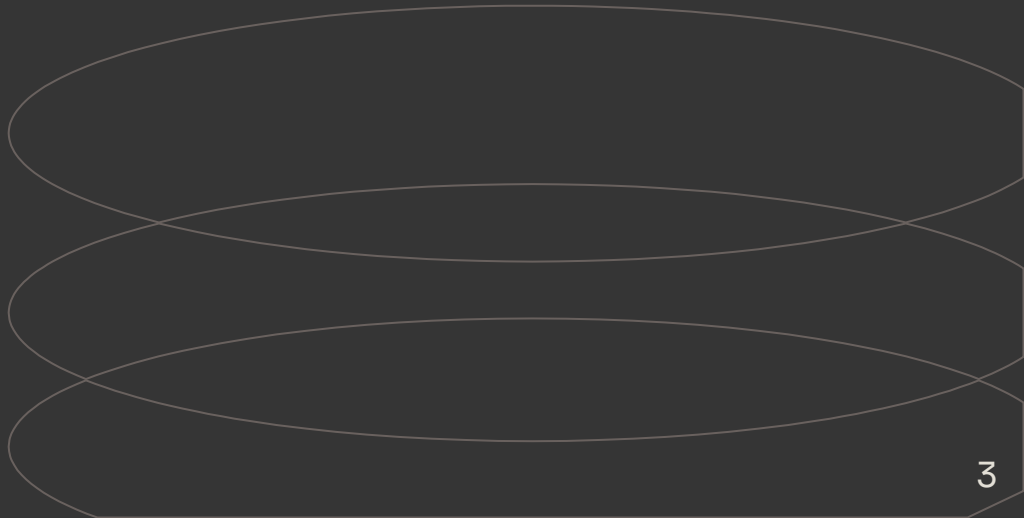
School: University of Michigan

Major: Physics

Minor: Complexity Science

Also: Co-Editor in Chief of the Michigan Daily

Introduction:
Science policy 101



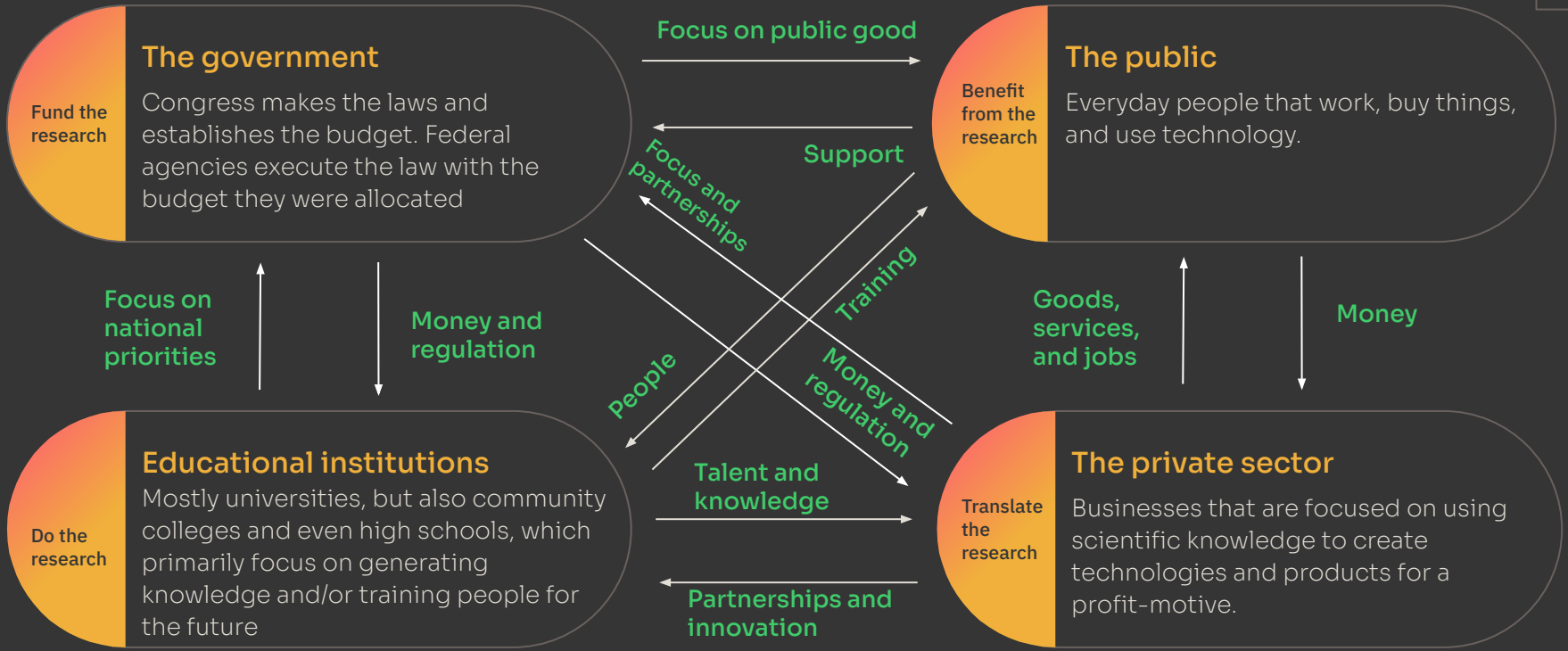
First, a crash course in science policy

To understand what kind of data we want and how we want to use it, we need to understand science policy

Science policy 101

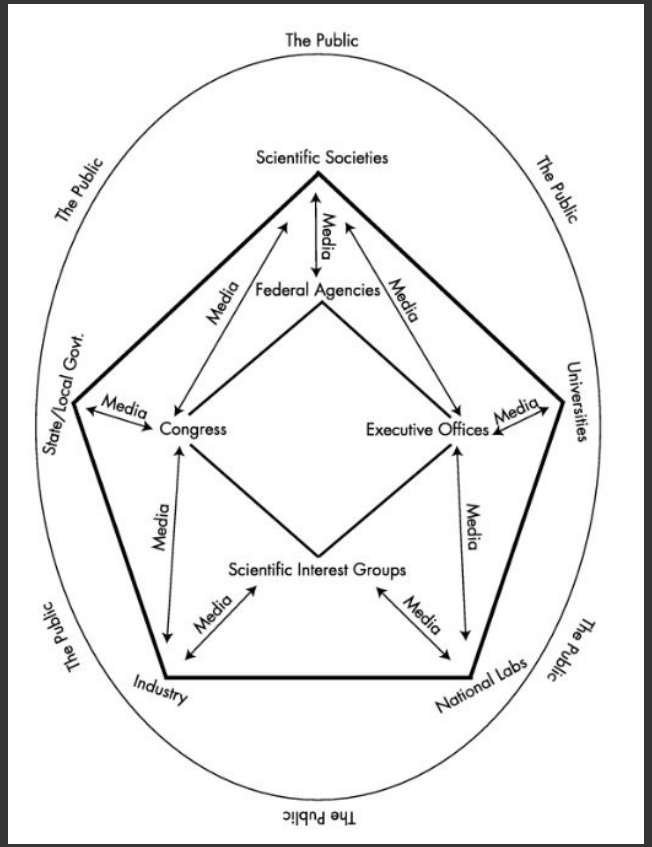
The science policy landscape is a collection of actors and incentives

The National Science Foundation is here!



Science policy 101

The science policy landscape is a collection of actors and incentives



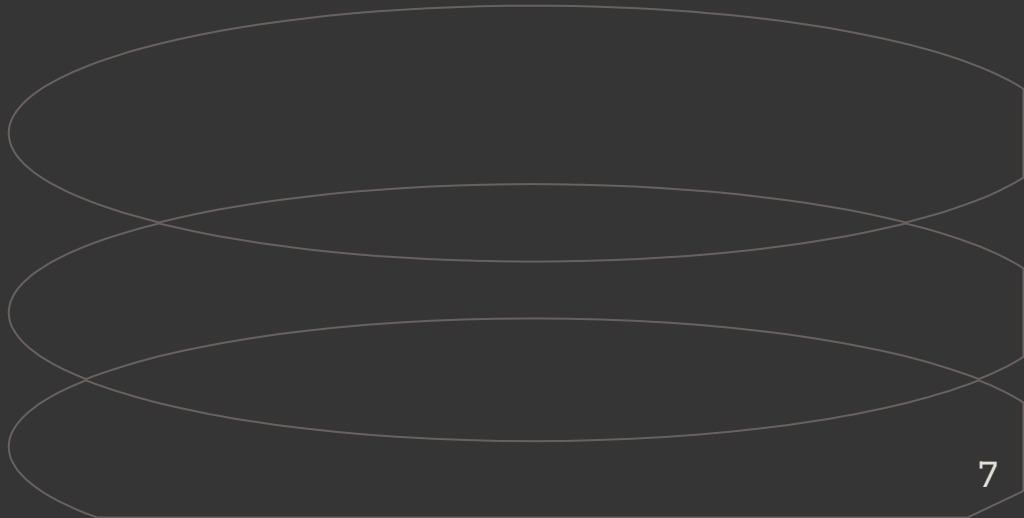
Note: This is...simplified

(but helpful!)

Neal, Smith, McCormick. Beyond Sputnik. 2008.

Data:

Necessary but not sufficient



We need data to:
(1) Inform policy
(2) Communicate impacts

The availability and quality of data directly impacts our ability to do these things

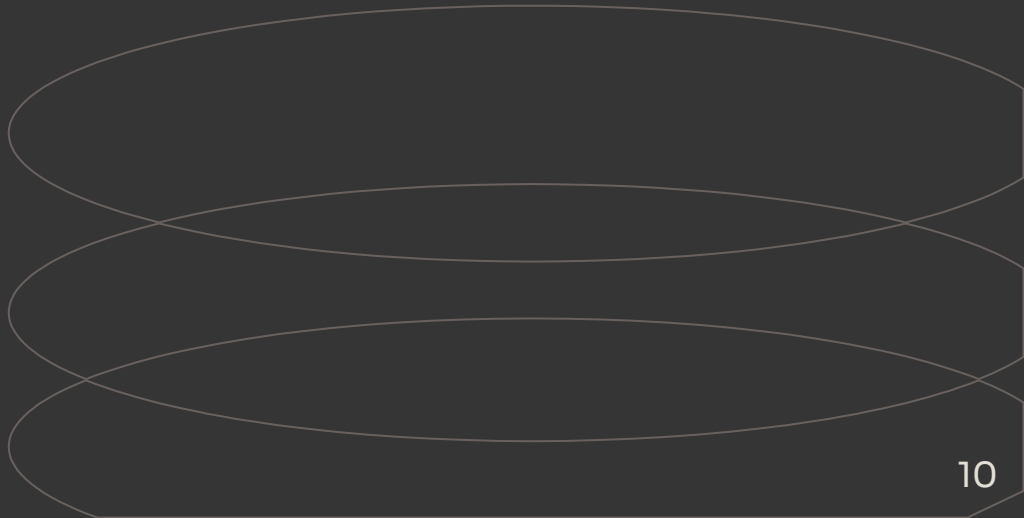
Necessary but not sufficient

Examples of policy-relevant questions that one might encounter that are somewhat difficult to answer given current available data infrastructure.

	Example question...	Downstream from....
1	What does NSF's return on investment look like for AI company seed-funding?	How good is the NSF at translating scientific ideas into economic benefit in critical tech areas?
2	How well is Michigan doing in science and engineering compared to Ohio?	What states are ahead in science and engineering and which are behind?
3	What proportion of highly cited AI- and quantum-related articles does NSF fund?	Is NSF an outsized player when it comes to generating new knowledge in critical tech areas?

Sets:

A python Excels in the library



My work

For more robust and dynamic data usage to fully capture the S&E landscape, we need more flexible data infrastructure

A python Excels in the library

Three data tools to answer three questions

NSF Award Search

Dataset: NSF Public Awards Database

What is it made from:

Python code and downloaded award files

What can it do:

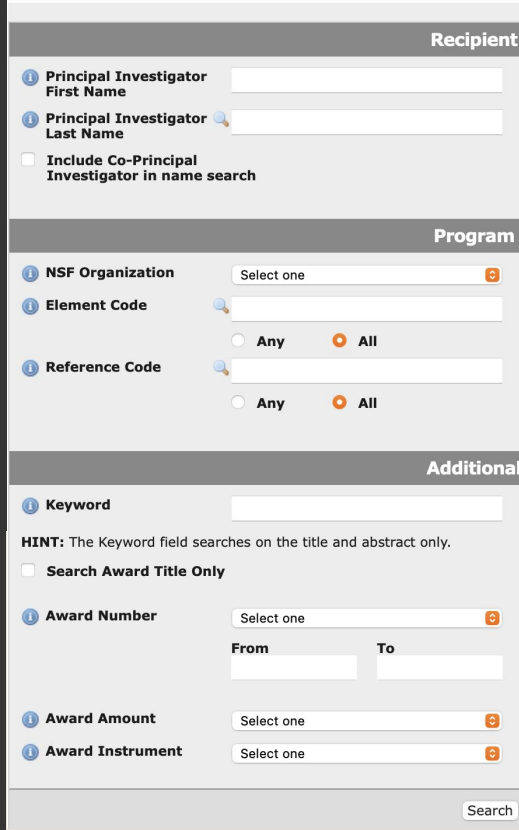
Filter all NSF awards by whatever characteristic we want (keyword, institution, PI…) and extract whatever information we want from them

What does ROI look like for NSF AI-related seed funding?

1. Filter awards by company seed funding program codes and keywords that define AI-related awards
2. Extract company names and award amounts
3. Match the company names with an outside database like pitchbook to find follow-on funding amounts. Or, search the web!

What individual award files look like


```
"awd_id": "2419989",
"agcy_id": "NSF",
"tran_type": "Grant",
"awd_istr_txt": "Continuing Grant",
"awd_titl_txt": "Collaborative Research:
"cfda_num": "47.050",
"org_code": "06030000",
"po_phone": "7032927944",
"po_email": "cshaw@nsf.gov",
"po_sign_block_name": "Colin A. Shaw",
"awd_eff_date": "2025-01-15",
"awd_exp_date": "2027-12-31",
"tot_intn_awd_amt": 483789.0,
"awd_amount": 173614.0,
"awd_min_amd_letter_date": "2025-01-13",
"awd_max_amd_letter_date": "2025-01-13",
"awd_abstract_narration": "Subduction zon
revealed more complex deformation events t
can influence the occurrence of large eart
```

A snippet of the NSF's current "advanced award search." It is limited by its keyword search and amount of awards it can return.

A python Excels in the library | Three data tools to answer three questions

Index Excel Calculator

Dataset: NCSES State Indicators

What is it made from:

Excel formulas and NCSES state-level data

What can it do:

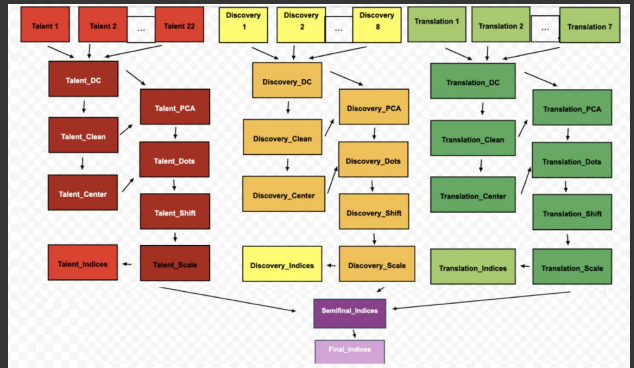
Take in statistical data for a given state and output an index. Several options exist for weighting indicators and normalizing and can be inputted with minimal tweaks

How well is Michigan doing in science and engineering, compared to Ohio

1. Find the Index for Michigan and Ohio and compare!

All credit goes to Andrew Czeidinski for the conceptualization and creation of the index. All I did was make an excel sheet that could (potentially) calculate it

Sheet diagram and example excel function




```
=LET( rawVal, INDEX([THEME]_Clean!$A:$Z, ROW(), COLUMN()), posInBlock, MOD(ROW()-1, 52), blockStartRow, ROW() - posInBlock, blockRange, INDEX([THEME]_Clean!$A:$Z, blockStartRow, COLUMN()):INDEX([THEME]_Clean!$A:$Z, blockStartRow + 51, COLUMN()), z, (rawVal - AVERAGE(blockRange)) / STDEV(blockRange), IF(ISNUMBER(z), z, 0) )
```

State Indicator
Fourth Grade Mathematics Performance
Fourth Grade Mathematics Proficiency
Fourth Grade Science Performance
Fourth Grade Science Proficiency
Eighth Grade Mathematics Performance
Eighth Grade Mathematics Proficiency
Eighth Grade Science Performance
Eighth Grade Science Proficiency
Individuals with High School or Higher-Level Degree among 25–44-Year-Old Population
Associate's Degrees in Science and Engineering (S&E) Conferred per 1,000 Individuals 18–24 Years Old
Associate's Degrees in Technology Conferred per 1,000 Individuals 18–24 Years Old
Bachelor's Degrees Conferred per 1,000 Individuals 18–24 Years Old
Bachelor's Degrees in Science and Engineering (S&E) Conferred per 1,000 Individuals 18–24 Years Old
Science and Engineering (S&E) Degrees as a Percentage of Higher Education Degrees Conferred
Science, Engineering, and Health (SEH) Graduate Students per 1,000 Individuals 25–34 Years Old
Advanced Science and Engineering (S&E) Degrees as a Percentage of S&E Degrees Conferred
Science and Engineering (S&E) Doctoral Degrees as a Percentage of S&E Degrees Conferred
Average Undergraduate Charge at Public 4-Year Institutions
Average Undergraduate Charge at Public 4-Year Institutions as a Percentage of Disposable Personal Income
Appropriations of State Tax Funds for Higher Education as a Percentage of Gross Domestic Product
State Expenditures on Student Aid per Undergraduate Student
State Support for Higher Education per Full-Time Equivalent Student
Postsecondary Degree Holders among Individuals 25–44 Years Old
Bachelor's Degree Holders among Individuals 25–44 Years Old
Bachelor's Degree Holders in the Labor Force
Individuals in Science and Engineering Occupations as a Percentage of All Occupations
Employed Science, Engineering, and Health (SEH) Doctorate Holders as a Percentage of the Workforce
Life Scientists as a Percentage of All Occupations
Computer and Mathematical Scientists as a Percentage of All Occupations
Physical Scientists as a Percentage of All Occupations

Snippet of some of the current state indicators. There are 63 in total — A lot!

The index would allow us to draw state level insights across all indicators

A python Excels in the library

Three data tools to answer three questions

Funding proportions search

Dataset: OpenAlex dataset

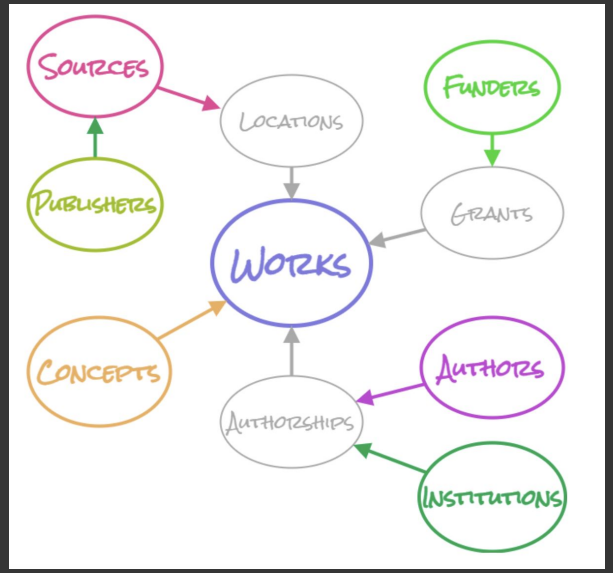
What is it made from:

Python code that queries the OpenAlex API

What can it do:

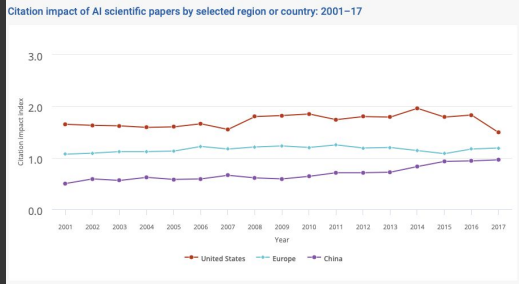
Queries the OpenAlex API to filter through articles and finds the funders

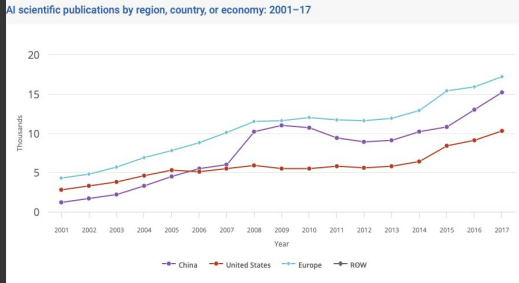
OpenAlex framework diagram



What proportion of highly cited US AI- and Quantum-related articles does the NSF fund?

1. Filter for articles with US authors, in the subfield AI or with “quantum” topics, in the top 1% of cited articles for its subfield
2. Group by funder and calculate proportions of articles in the subset funded by NSF!

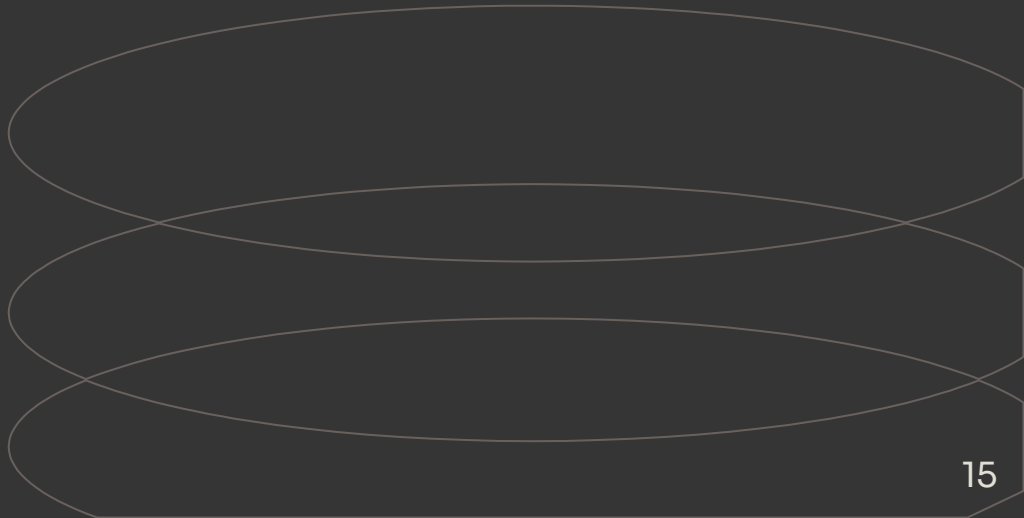




Examples of currently available publications data from Scopus. The top is citation impact by country over time and the bottom is proportion of AI publications by country over time.

While the Scopus database has “higher quality” articles, OpenAlex is much bigger and contains more information about funding.

Test:
A dual at high noon



The setup

On Thursday, multiple conversations spur demand for an NSF AI and Quantum Factsheet

The argument: “NSF is key to U.S. leadership in AI and quantum.”

We need a rough draft by Monday so that we can put it through the editing and design process over the next week and release it by the Board meeting on the next Wednesday.

NSF Investments Are Key to US Leadership in AI & Quantum

National Science Foundation: *A Change Agent for the Change Agents*

Discoveries in the fields of Artificial Intelligence (AI) and Quantum science hold the promise to catalyze economic growth and transform the way we live. **The NSF is an indispensable player in the AI and Quantum ecosystem, seeding promising ideas and companies with the funds to create breakthroughs and take them to market.**

How has the NSF already impacted the AI and Quantum enterprise?

- **Funding early research in stochastic modeling and neural networks**, which are the basis for many of today's AI tools like ChatGPT, and paving the way for industry's now annual investment of tens of billions of dollars toward AI research ad development.
- **Supporting early quantum algorithm and qubit hardware research**, demonstrating the capabilities and feasibilities of quantum computing, and igniting a wave of interest by industry leaders to build upon this NSF-funded research.
- **Accelerating the transition from knowledge into economic competitiveness** by providing entrepreneurship training to researchers and seed funding to startups creating game-changing technologies. The NSF's Technology, Innovation, and Partnerships directorate heads these programs for strategic coordination matching agency priorities.[1]

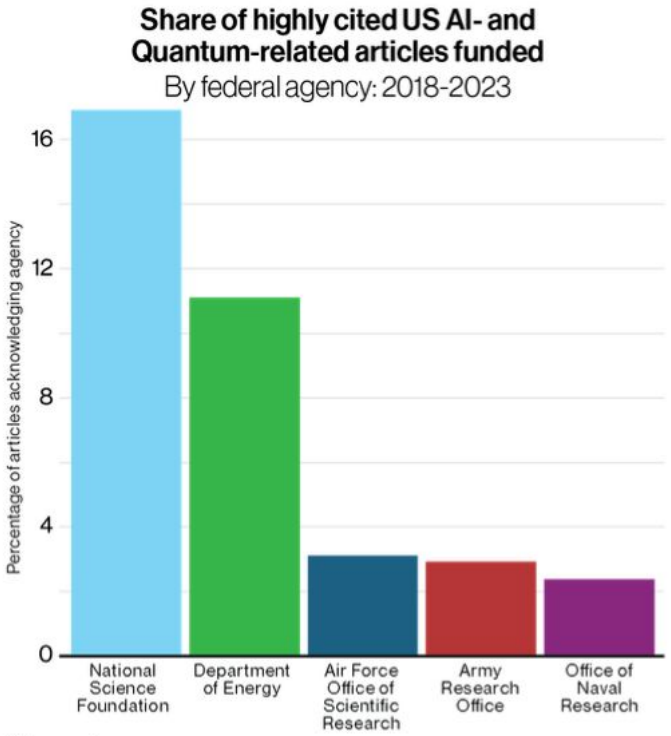


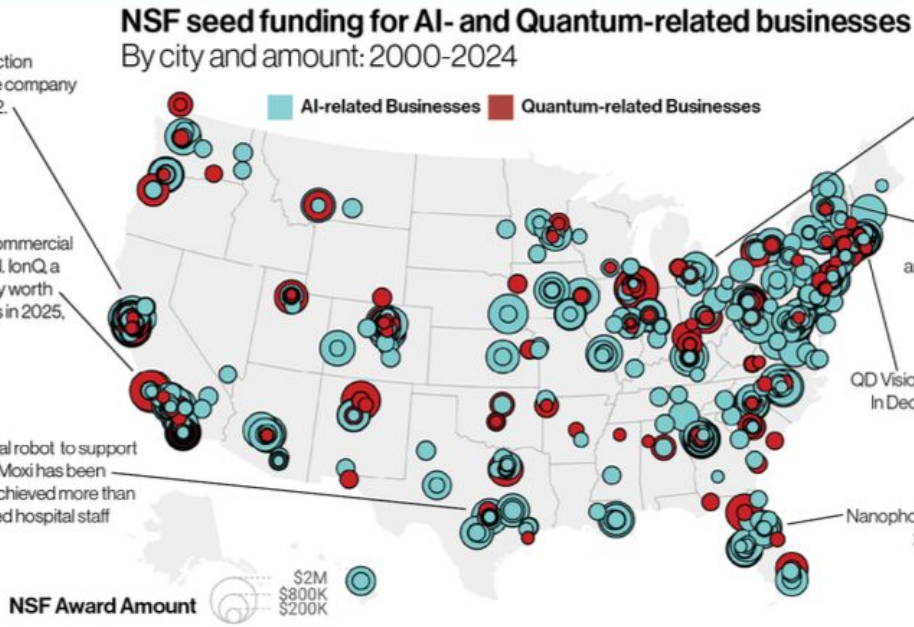
Figure 1

Since the 2022 fiscal year alone, the NSF has invested more than $289M in seed funding for almost 600 AI- and Quantum-related companies, setting the stage for venture capital (VC) funding and paradigm shifts in a wide range of fields. Here's a look at some of the companies and fields the NSF has funded over the years:

Construction
OpenSpace, Inc. provides complete, AI-powered visual records of construction projects for project management. The company was recently valued at $902M in 2022.
NSF 2018: 0.975M
VC in 2022: $157M

Quantum Computing and Networking
Qubitekk, Inc. powered the first U.S. commercial quantum network in Chattanooga, TN. IonQ, a leading quantum computing company worth over $13B, acquired Qubitekk's assets in 2025, including its 118 patents.
NSF 2016: $0.225M
VC in 2019: $2.1M

Healthcare
Diligent Robotics built the Moxi hospital robot to support clinical staff teams on logistical tasks. Moxi has been deployed in more than 30 hospitals, achieved more than 1 million in-hospital deliveries and saved hospital staff more than 600,000 hours.
NSF 2017: $0.725M
VC in 2022: $30M



Banking
Clinc, Inc. is an Ann Arbor, MI based firm that builds AI-powered chatbots for large banks, enabling them to effectively handle over 95% of inquiries. More than 5 million unique banking customers use their services.
NSF 2017: $0.975M
VC in 2019: $52M

Electronics
Green Mountain Semiconductor Inc. is a design-services and IP company specializing in advanced circuit and memory architectures for applications in AI, aerospace, defense, and space.
NSF 2018: $0.9865M
Estimated Annual Rev: $4.2M

Next-Gen Displays
QD Vision develops Quantum Dot Films to enhance LCD TV color ranges. In December 2016, Samsung acquired QD Vision's intellectual property, including around 250 patents for $70M.
NSF 2011: $0.15M
VC in 2015: $50.9M

Nanophotonica, Inc. develops quantum dot materials for QLED displays. In 2019, it closed a series A funding round led by Samsung Ventures.
NSF 2014: $0.893M
VC in 2019: $3.5M

Created by Datawrapper

Figure 2

1. NSF TIP Directorate, America's Seed Fund Program, I-Corps Program, Regional Innovation Engines

Figure 1: OpenAlex Database, July 2025. Highly cited articles are defined as those with a field-weighted citation index in the top 1%. AI-related articles given by "Artificial Intelligence" subfield. Quantum-related articles given by topics with the keyword "Quantum." AI and Quantum articles taken together due to substantial overlap publications that are AI-related and Quantum-related. Filtered for articles with at least one U.S. author.

Figure 2: NSF Awards Database, July 2025. "Seed funding" given by Element codes: 5371, 5373, 1505, 1591. AI- and Quantum-related awards were identified using keywords "Artificial Intelligence" and "Quantum.". Other data gathered from online press releases and other public information. Total amount invested and number of companies is a lower limit due to search method. Created by Datawrapper.

National Science Board
NationalScienceBrd@nsf.gov | 703.292.7000

NSB-2025-15
https://www.nsf.gov/nsb/publications/2025/NSFAIQ.pdf

How has NSF reached such a pivotal position?

Through strategic investment in AI and Quantum talent and research.

AI and Quantum Information Science Research and Development funding
By federal agency: FY2023 Actual

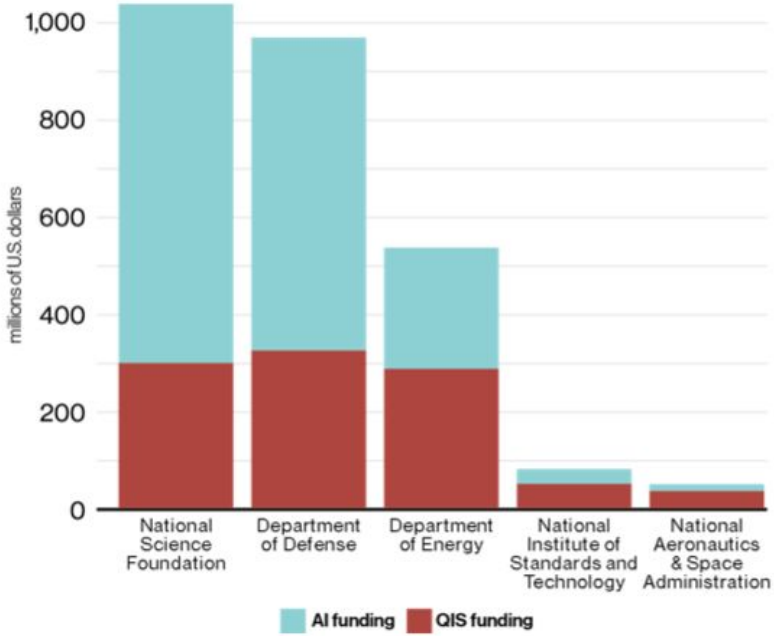


Figure 3

Primary source of federal support for full-time graduate students
By broad field: 2023

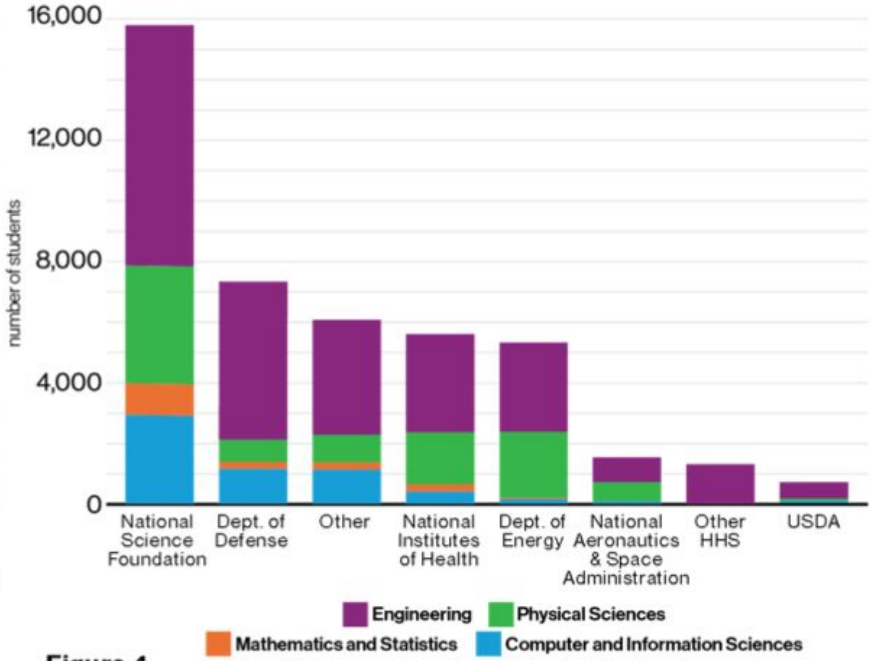


Figure 4

NSF is unique among federal agencies in its approach to supporting emerging technologies like AI and Quantum Science, funding both talent and R&D in concert. From 2000-2022, NSF has funded 6900+ PhD dissertations in Quantum and AI and propelled the careers of field leaders like:[2]

Geoffery Hinton 2025 Nobel Prize Winner in Physics for early work in neural networks. His research was funded by NSF in 1986.

Yann LeCun Chief Scientist at Meta AI. Received over $750k in NSF research funding in the mid-to-late 2000s.

Peter Shor Demonstrated capability of quantum computers to break RSA encryption. Received multiple NSF awards for similar research.

Prineha Narang Founder and CTO of Aliro, a post-quantum cryptography company with over $9M in VC funding. Graduate and early career work supported by NSF.

NSF investments in STEM education and workforce training like these come from research-oriented and education-oriented programs alike. NSF-funded talent is therefore closely connected to NSF-funded basic research and facilities, creating a self-reinforcing network of STEM collaboration and advancements. Part of this holistic investment is a focus on building broad and deep collaboration networks and innovation ecosystems.

Ways the NSF is advancing U.S. leadership in AI and Quantum:[3]

- **Training the workforce**. The Advanced Technology Education (ATE) program provides hundreds of two-year colleges with grants to modernize their courses, purchase up-to-date equipment, and retrain instructors so the U.S. has a strong workforce of technicians trained for the future.
- **Doing the deep science.** The NSF's 5 Quantum Leap Challenge Institutes are multi-campus collaborations focused on creating quantum breakthroughs through collaboration, education, and innovation. And, with $540M in direct investments and partnerships across federal agencies and the private sector, the NSF's 27 AI research institutes connect over 500 institutions together to advance AI research.
- **Creating economic growth.** Regional Innovation Engines (RIEs) serve to spur regional and national economic growth. They form regional coalitions of institutions, researchers, and companies that advance critical technologies and address pressing national challenges.

NSF programs driving U.S. leadership in AI and Quantum
By city and type

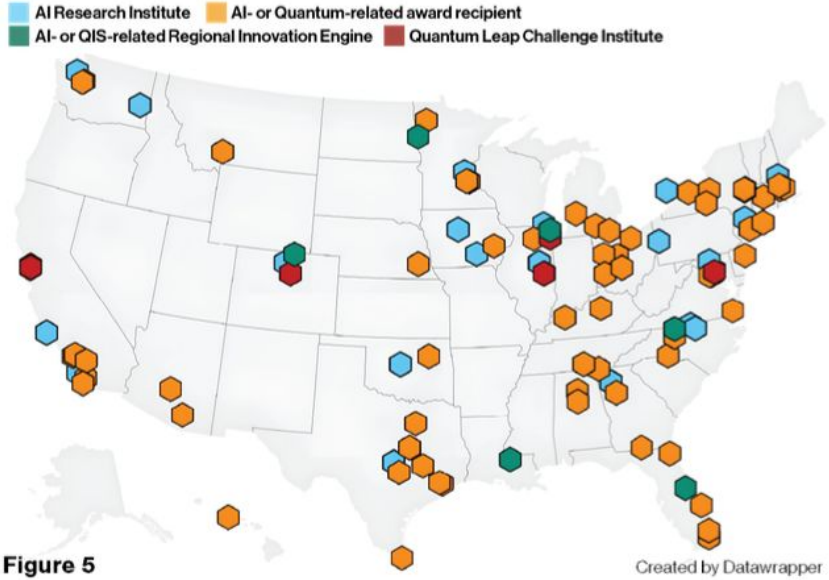
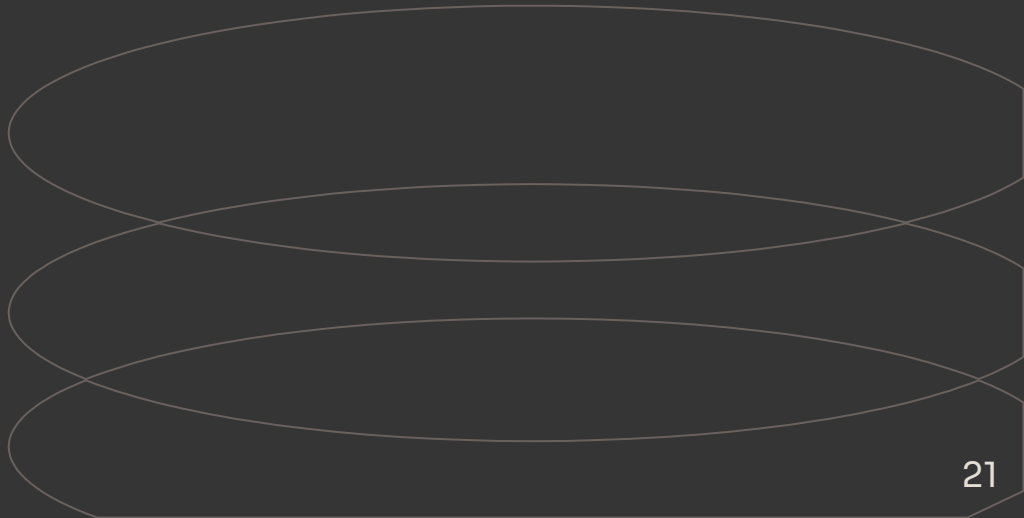


Figure 5

Forward:
Summarizing proposals for the future



General proposals for the future

Finding the right data and using it the right way

1. **Making the data tools more robust**

2. **Using ML to organize databases**

3. **Answering other pressing questions**

Special thanks to the following people

Acknowledgements

1

AIP and SPS

For their assistance with the entire internship — the logistics, placement, and programming — and for giving me the opportunity to be an SPS intern.

2

NSBO Staff

For their support, mentorship, and patience as I worked across these multiple projects. And for making the NSB such an amazing place to work and learn.

3

Amanda Vernon

For everything. I am truly sorry for every future mentor have in life because of the insanely high bar she has set.

If anything jumps to mind later, feel free to email me at zhane@umich.edu

Questions? Ask away!


Thank you!