



Teaching AI How Science Actually Works

How block-grant labs can generate the real-world data AI needs to do science | Ben Reinhardt

Teaching AI How Science Actually Works

How block-grant labs can generate the real-world data AI needs to do science | Ben Reinhardt

This essay is part of [The Launch Sequence](#), a collection of concrete, ambitious ideas to accelerate AI for science and security.

Summary

Current AI for science consists of specialized tools built on curated datasets, but much of science is hands-on experimentation and unrecorded tacit knowledge. Science leadership will belong to whoever figures out how to use AI for this messy part of science. For AI scientists to become truly capable, they will need multimodal datasets capturing that tacit knowledge.

We propose an ambitious effort to generate and use the multimodal data necessary to unlock the full potential of AI for science. This program would create “Unstructured Data Generation Labs” that simultaneously conduct breakthrough research, comprehensively record everything from bodycam videos to keystrokes, and use that data to make the whole process more productive.

Each organization will focus on equipment-defined domains like biotechnology, advanced materials and manufacturing, or micro/nanotechnology. Institutional block grants, expert-review-based tranches, and default sunseting will give the labs the freedom to pursue unexpected directions while maintaining oversight. A security organization will mitigate potential risks from malicious actors misusing the data and models.

The program will cost \$2 billion over eight years. Implementation will have three phases:

- **Phase 1:** \$200 million over two years for 20 pilot organizations.
- **Phase 2:** Down-selection to 5 organizations that receive \$50 million per year over three years to generate initial datasets and breakthroughs.
- **Phase 3:** An additional \$50 million per year over three years to scale up datasets and transfer technology to the broader innovation ecosystem.

This approach could accelerate research productivity by 10-100× in targeted fields while maintaining safety through controlled access and comprehensive oversight.

Motivation

AI-powered science represents civilization-scale opportunity. The dream of AI for science is straightforward: a legion of graduate students, technicians, and potentially senior researchers who could work 24/7 anywhere in the world, don't get bored or distracted, and don't require years of training before they are productive. Functional AI scientists could reverse decades of declining research productivity, accelerate the discovery of wonder materials and life-saving drugs, and unlock transformative energy technologies.

Why we don't have AI scientists: The tacit knowledge gap

Current AI for science like AlphaFold or AI-designed catalysts have demonstrated incredible potential, but they share a key limitation: models don't actually "know" how to do science.

Consider what it actually takes to develop new materials. Say you want AI to accelerate scalable carbon nanotube synthesis for next-generation electronics. The published literature describes successful synthesis conditions: "heat carbon feedstock to 800°C in CVD reactor with nickel catalyst." But actually making this work requires knowing that the reactor needs 3 hours to thermally equilibrate, that

contamination from the previous run affects yield, that "nickel catalyst" means specific particle sizes prepared in a particular way, and that the carbon feedstock quality varies by supplier batch. When synthesis fails — which it does constantly — success depends on recognizing whether the problem is temperature uniformity, gas flow turbulence, catalyst poisoning, or substrate preparation. This troubleshooting process involves adjusting dozens of parameters based on visual cues, equipment sounds, intermediate measurements, and hard-won intuition about what "normal" looks like.

The same pattern repeats across scientific domains. Cell culture protocols omit that different incubator positions have temperature gradients affecting growth rates. Quantum optics papers don't mention that laser alignment drifts with building vibrations throughout the day. Materials characterization requires knowing which sample preparation artifacts look like real signals. Even computational work depends on physical intuition — molecular dynamics simulations need parameters derived from experimental observations, and validating results requires understanding how real molecules actually behave.

This tacit knowledge gap explains why most AI for science remains narrowly specialized. Current approaches work when the relevant knowledge can be captured in clean datasets — protein structures, chemical properties, literature relationships. But they fail when success depends on integrating information across multiple scales, modalities, and domains of expertise. Training AI only on papers and databases is like trying to learn surgery from textbooks without ever watching an operation or handling instruments.

Most work in science doesn't happen entirely in computers or make for well-structured datasets. Research papers describe successful outcomes, not the full iterative process that creates them. So much of science is the prosaic work to adjust equipment, troubleshoot a failed experiment, or notice "huh, that's funny." Even the work that does happen in computers (simulations, trying to make sense of sensor outputs, or designing experiments) needs to be grounded in physical reality.

AlphaFold leveraged a meticulously gathered database of protein structures. Most approaches to AI for science follow this pattern: collect a meticulously curated dataset either by hand or with "self-driving labs" and then train a specialized model

on it, whether that dataset is papers, potential materials, or druggable targets. Current “AI-for-science” approaches resemble AI before Large Language Models (LLMs): models with impressive, but narrow applications, trained on curated datasets of text or images.

However, the massive breakthrough that enabled the current generation of AI was not meticulously gathered datasets, but the ability to train on the entire internet. The internet is basically a compilation of what humanity does digitally. As a result, AI has become incredibly good at things that live entirely in computers — writing, coding, generating video. But these capabilities haven’t translated to broadly useful AI for science.

Why existing institutions won’t solve this problem

To produce the unstructured data needed to actually make AI broadly useful for science, we need new labs. This is because existing research institutions have fundamental organizational and incentive barriers that will prevent them from succeeding at this mission. Universities are structured around independent labs and training graduate students; this structure is good for traditional ways of doing science, but will run counter to the organization-wide coordination needed to create and use unstructured data. Both national labs and universities have become incredibly bureaucratic — whereas building and running these data generation labs successfully requires the ability to move quickly and make non-consensus decisions. Moreover, this work will require innovative team organization, new workflows, and built-from-scratch equipment that will run counter to how existing organizations do things. It would be just as hard, if not harder, to retrofit existing equipment and overhaul operating procedures, compared to standing up entirely new labs.

Corporations are unlikely to do this work except perhaps in highly profitable niches like health-focused biotechnology. Large corporations have already gutted their R&D departments, shortened their timescales, and offloaded much of their innovation to startups. Startups experience extreme pressure to specialize in profitable niches and build products rather than do broadly applicable research work. And, while the profit-seeking is a useful feedback loop, it will also keep

companies from sharing their outputs with the larger scientific community that can leverage them.

Solution

To unlock AI that is truly useful in many areas of science, we need to create a new kind of institution: labs that simultaneously do cutting-edge science, collect data on how that science actually happens, and use that data to train models that they use to do the science even better.

These “unstructured data labs” need the following ingredients:

- Scientists who are doing serious work to create actual discoveries and inventions.
- Heavily “instrumented” labs to collect data on everything that is going on — from bodycams, to logging every computer keystroke and instruction sent to every machine.
- Separate teams devoted to doing the actual research, data collection and AI tools, diffusing data and knowledge out of the lab, and dedicated technical support for all of the other groups. It’s critical that each of these teams have similar status and resourcing.
- Carefully constructed incentives to get the best people to do actually useful work.

Scope

The labs must do three things:

1. Serious research work
2. Collect data on all aspects of how that work happens
3. Train productivity-enhancing AI models on that data

This is a broad scope for a single organization, but it's important that it all happen under one roof for two reasons:

1. If you don't collect data on serious work, the resulting models won't actually be able to do serious work.
2. It won't be obvious at first what kind of data or metadata will be important for creating useful tools so if you don't do all three, you could end up collecting data that is useless for creating useful models.

Certain scientific fields will advance our ability to manipulate the physical world and determine US competitiveness over the next decades. We should scope labs around the equipment and techniques that define these areas, rather than placing bets on any narrow topic du jour. This will allow them to do useful work over their full lifespan, rather than chasing a particular goal that could fall out of vogue. Initial fields may include:

- **Biotechnology and biomanufacturing:** Cell culture, genetic engineering, protein production, and therapeutic development.
- **Quantum systems and photonics:** Laser spectroscopy, quantum sensing, optical component design, and precision measurement systems.
- **Advanced materials and manufacturing:** Synthesis of novel composites, advanced manufacturing processes, materials characterization, and scalable production methods.
- **Micro- and nanotechnology:** Cleanroom fabrication, electron beam lithography, micro-electro-mechanical systems devices, and nanoscale characterization techniques.
- **Systems biology and ecology:** raising and analyzing animals, plants, fungi, along with the fieldwork to discover new secrets of nature.

These institutions need unconventional funding mechanisms to get the best talent doing ambitious work and enable meta-experiments on data collection and AI tools. Traditional project-based grants and line-item budgeting would constrain the iterative, constantly changing work to build functional systems, scare off the best talent, and push the organization towards showmanship instead of real results.

Instead, funding for these labs should come from a combination of institutional block grants and ongoing contracts with industry partners and government agencies who want to use both the labs' research outputs and data. The former could be implemented as an Other Transaction Authority (OTA), potentially using the recently-proposed [X-Labs framework](#).

This proposal will cost \$2 billion over the course of 8 years — a low price for the potential to unlock broadly-useful AI for science.

Program oversight and governance

Oversight should happen through comprehensive review-based tranches at the end of years two and five.

Because this type of organization is so new, it will be hard to successfully select the exact right proposals up front. Instead, the initiative should be started as a competitive pilot program with down-selection after the first two years. Those selected organizations should then be subject to comprehensive review three years later to decide whether to continue their funding.

The reviews should require each lab to be retroactively judged by experts on what they've accomplished at the predetermined intervals. This approach is different from milestone-based funding because the nature of research definitionally involves uncertain outcomes and timelines. However there should be a broad agreement on what "good" looks like in order to move onto each new tranche. The [MRC Laboratory of Molecular Biology](#) used this approach to win 12 Nobel prizes.

Furthermore, the labs should sunset after eight years by default. To make organizations with a wide mandate politically palatable, avoid mission creep, and mitigate the tendency for organizations' purpose to become nothing more than self-preservation, the organizations should sunset by default after a fixed period of time. This time period should be more than five years, but less than ten.

Timeline

Phase 0

- Appoint a program director
- Put out call for proposals

Phase 1: 2 Years, \$200 million

- Award 20 organizations \$10 million each to demonstrate proof-of-concept.

Success looks like: functional instrumentation prototypes, initial datasets, evidence of research productivity under heavy monitoring. While many organizations will “fail,” their work will provide valuable negative results and they can still go on to raise private funding to continue their work.

Phase 2: 3 Years, \$900 million

- Down-select to five organizations based on comprehensive expert review.
- Full \$50 million/year funding for each of the five organizations.

Success looks like: Initial research breakthroughs internal to the labs that wouldn’t have happened without the AI and useful datasets.

Phase 3: 3 Years, \$900 million

- Re-authorized organizations continue research work and data collection with additional focus on diffusing breakthroughs, data, and techniques into the broader innovation ecosystem.
- At the end of this phase, the organizations should shut down by default. They could figure out an ongoing business model as an industry consortium, be acquired by a corporation or specific agency, etc.

Success looks like: Robust datasets, useful AI science tools diffused into the US innovation ecosystem, research breakthroughs external to the labs.

Risk mitigation

A big reason to fund this work with public resources is to do it in a way that adds security while broadening access: instead of leaving unstructured data generation to happen at some VC-backed private AI firm for its own use, this proposal can:

1. Make the data a public good by providing free or subsidized access to legitimate and responsible users.
2. Make it actually hard for terrorists or irresponsible users to access the data.

Like any effort to increase AI capabilities and enhance scientific productivity, unstructured data generation labs may raise several concerns about privacy and risks. Best practices in AI security and risk mitigation are moving quickly. Instead of a static set of policies that will likely be obsolete by the time these labs are built, a separate security organization should create, update and implement best practices for minimizing the chances that the labs or their outputs will aid malicious actors.

Some examples of security policies might include the following, though the exact policies should be left up to the security organization:

- Excluding the highest-risk research areas like [gain-of-function virology](#) or weapons-relevant chemistry.
- Limiting raw data access and more powerful models to responsible users who have been thoroughly vetted.
- For privacy reasons, all datasets should be scrubbed of individually identifiable information like faces, voices, etc.

The security function could be done either by a new organization that is spun up as part of this effort or be contracted out to existing organizations that already have experience with risk mitigation and security for powerful models.

However, there is no airtight way to increase general-purpose scientific capabilities without increasing the ability for people to do bad things with it.

Recommended actions

Congressional authorization

- **Authorize \$2 billion over 8 years** through NDAA or America COMPETES reauthorization for “Unstructured Data Generation Labs for AI Science.”
- **Establish a joint program office** spanning DOE, DOD, NSF, and NIH with streamlined oversight authority.

Appropriations

- **Phase 1:** \$200 million over 2 years for competitive pilot program (20 organizations × \$10 million each)
- **Phase 2:** \$900 million over 3 years for down-selected institutes (5 organizations × \$60 million/year each)
- **Phase 3:** \$900 million over 3 years for continued operations and technology diffusion

Implementation mechanisms

- **Enable Other Transaction Authority (OTA)** for block grant funding without traditional line-item oversight.
- **Authorize industry cost-sharing agreements** allowing private partners to contribute 25-50% funding in exchange for preferential data access. This cost sharing will enable work to start more quickly and make sure that the work is tied more closely to outcomes that are actually useful.
- **Establish expedited security clearance process** for researchers working on dual-use data collection systems.
- **Create statutory exemption** from standard federal procurement rules to enable rapid hiring of top talent at competitive salaries.

Oversight structure

- **Appoint Senate-confirmed program director** within 90 days of authorization.

- **Mandate comprehensive reviews** at years 2 and 5 by independent expert panels, with automatic sunset after 8 years unless explicitly reauthorized.
- **Mandate chemical, biological, radiological, and nuclear (CBRN) capability evaluations** before releasing any AI models or datasets to external users.

Further resources

- James Phillips, "[Ideas on scaling technoscience](#)," n.d.

On block grants and comprehensive review.

- Caleb Watney, "[Launching X-Labs for Transformative Science Funding](#)," n.d.
- Understanding AI, "[I got fooled by AI-for-science hype](#)," n.d.

On the limitations of current approaches in AI-for-science.