PRIORITIZING INTERNATIONAL AI RESEARCH, NOT REGULATION

Kevin Frazier*

September 2024

Given the nascent stage of artificial intelligence, regulation should give way to research. An international research entity carries the greatest odds of detecting AI risks and identifying its benefits.

Rapid advances in artificial intelligence (AI) have sparked a wave of regulatory efforts.[1] Lawmakers at the state and federal levels have made innovation the primary focus of AI regulation.[2] Their secondary priority is mitigating risks to reduce the odds of short- and long-term dangers posed by AI.[3] The

* Kevin Frazier is an assistant professor at St. Thomas University College of Law and co-director of the Center for Law & AI Risk. He is also a Senior Research Fellow at the Constitutional Studies Program at the University of Texas at Austin. Participants in the AI Futures Fellowship and VISTA AI Institute Law and Policy Workshop provided useful feedback on earlier drafts of this paper. Alan Rozenshtein's thorough edits and thoughtful suggestions strengthened the paper as well. The editors of the *Washington Journal of Law, Technology & Arts* offered insightful commentary on a version of this paper published in volume 13 of their journal.

[1] See, e.g., "Artificial Intelligence Legislation Tracker," Brennan Center (May 7, 2024), https://perma.cc/WW34-LFBP (listing legislation to regulate AI).

[2] See Madison Alder, "Schumer Says AI Roadmap Coming Soon From Senate Working Group, FedScoop (May 8, 2024) (quoting Majority Leader Chuck Schumer as touting innovation as the "North Star" of regulatory efforts); Gaeme Hanna, "Two US States Seize the Initiative to Regulate AI," ReadWrite (May 16, 2024).

[3] By way of example, the order of policy topics outlined in the Bipartisan Senate AI Working Group's Roadmap for Artificial Intelligence Policy suggests a prioritization of innovation and AI development more generally. Majority Leader Chuck Schumer et al., "A Roadmap for Artificial Intelligence Policy in the U.S. Senate," U.S. Senate (May 21, 2024).

Roadmap for Artificial Intelligence Policy released by the Senate's AI Working Group illustrates the focus on regulation in support of innovation, rather than on investments in AI risk research.

Intended by its authors, including Majority Leader Chuck Schumer (D-N.Y.), to speed up and coordinate AI governance proposals, the Roadmap has received broad attention from legislators, the press, and, by extension, the public. The core themes of the Roadmap provide a strong indicator of how lawmakers currently perceive the regulatory challenges posed by AI. The first section of the Roadmap is titled "Supporting U.S. Innovation in AI" and includes the working group's recommendation that at least $32 billion per year be spent specifically on AI innovation.[4] The penultimate section, "Safeguarding Against AI Risks," acknowledges that AI systems may present "a variety of risks" but offers far less specific and significant efforts to study those risks. The working group urges AI companies "to perform detailed testing and evaluation to understand the landscape of potential harms[.]" They also instruct relevant congressional committees to "[s]upport efforts related to the development of a capabilities-focused risk-based approach[.]"

A regulate-to-innovate mindset has also informed AI governance at the state level. For example, Connecticut Gov. Ned Lamont cited the need to support AI innovation when he successfully tanked a comprehensive AI bill by threatening to veto any such legislation.[5] Likewise, observers expect that Colorado's AI legislation will be watered down before it goes into effect to reduce the odds of it stifling innovation.[6] As discussed in more detail below, even AI research proposals at the state level have been framed more so as a means to foster innovation than to study and detect AI risks.

The European Union has also rushed to regulate AI, albeit with the aim of policing its use in sensitive contexts.[7] The EU AI Act, though, neglected to address the need for AI risk research. The United Kingdom, in contrast, recently announced research grants for experts to look into ways to shield society from AI risks.[8]

This paper challenges the regulate-first approach to AI governance. A better framework is to establish a research-regulation (R&R) cycle. This cycle starts with verifiable, reliable, and timely research on the risks posed by new technology. It then distills that information into actionable and understandable policy

[4] Sen. Chuck Schumer et al., "Driving U.S. Innovation in Artificial Intelligence," at 5 (May 2024).

[5] Sage Lazzaro, "Connecticut Swings Big on AI Regulation—and Strikes Out," *Fortune* (May 9, 2024).

[6] Matt Scherer, "Colorado's Artificial Intelligence Act Is a Step in the Right Direction. It Must Be Strengthened, Not Weakened," Center for Democracy & Technology (May 22, 2024), https://perma.cc/9CLV-BCR6.

[7] European Commission, AI Act (accessed May 22, 2024), https://perma.cc/857W-7GTX.

[8] U.K. Department for Science, Innovation, and Technology, "Tech Secretary Unveils £8.5 Million Research Funding Set to Break New Grounds in AI Safety Testing," Press Release (May 22, 2024), https://perma.cc/3M4W-KJSF.

recommendations. The resulting responsive regulation implements those recommendations and restarts the cycle by funding a new wave of research into unresolved questions.

Although each step of this cycle deserves close examination in the context of AI, this paper focuses on the first step: risk research. How best to conduct AI risk research in a way that fosters the research-regulation cycle has received inadequate attention from scholars and regulators alike. This paper fills that void through three inquiries. The first is whether subnational, national, or international bodies are best suited for the research in question. I conclude that an international body is superior. The second and third inquiries are what an international AI risk research initiative could learn from the European Organization for Nuclear Research (CERN) and the Intergovernmental Panel on Climate Change (IPCC), respectively.

CERN and the IPCC have received substantial attention as models for an international AI body.[9] A CERN for AI, on paper, would emulate CERN's approach to studying the origins of the universe: bringing experts to a common location to research the capabilities of advanced AI systems. This paper adds to the ongoing discussion of this proposal by diving into the unique historical conditions that led to CERN's creation as well as the distinguishing aspects of particle physics research. That close study of CERN's history and structure suggests that a CERN for AI is unlikely to work in practice.

An IPCC for AI, by comparison, is a more feasible approach to international AI risk research. Akin to the IPCC's regular assessment of the current state of climate science, an IPCC for AI would convene AI experts at fixed intervals to form a consensus understanding of the threats and capabilities of AI systems at that time. By virtue of requiring fewer resources and aligning with the distributed nature of AI research, an IPCC for AI seems more likely to succeed.

Neither the IPCC nor CERN is a perfect fit for AI risk research. That said, this paper concludes with high-level lessons an international AI risk research initiative should glean from the two organizations. Adherence to these lessons would increase the odds of international AI risk research getting underway and, per the R&R cycle, informing regulation.

Background on Risk Research

Research and regulation of an emerging technology are distinct endeavors. Pure research aims to understand the pros and cons of a technology. Regulation creates legal incentives to spread and augment those pros as well as incentives to stem the realization of those cons. The R&R cycle anticipates that research will precede regulation. Research into an emerging technology serves two purposes: first, identifying the proper regulatory responses and, second, justifying and orienting financial and political outlays for additional research.

[9] See Gary Marcus, "A CERN for AI and the Global Governance of AI," Substack (Oct. 26, 2023); Nicolas Miailhe, "Why We Need an Intergovernmental Panel for Artificial Intelligence," Our World (Dec. 21, 2018).

This cycle has played out in several contexts other than AI governance. Take, for example, auto safety standards. Insurers, local and state governments, and other stakeholders collect information on automobile accidents and send that information to the Insurance Institute for Highway Safety (IIHS) and the National Highway Safety Traffic Administration (NHTSA). Experts at the IIHS and NHTSA then develop crash tests based on those accident reports to pinpoint weaknesses and flaws in automobile designs.[10] These tests follow a publicly available protocol, and the results are made widely accessible in understandable formats.[11] In turn, regulators update safety standards as well as crash reporting requirements.[12] The next generation of cars are then subject to even more rigorousand sophisticated IIHS and NHTSA testing.

Yet, whereas regulation of an emerging technology is inevitable—after all, popular use of a technology will inevitably reveal its risks—risk research is not. The IIHS, for example, came about only *after* the dangers of cars had been made obvious.[13] For decades, discovery of the risks posed by cars was generally a matter of (bad) luck. Regulation was crafted in response to accident reports that often lacked critical information about the cause, severity, and preventability of accidents. Eventually, insurers realized that better safety information could improve their bottom line: More drivers using safer cars would lead to fewer claims. They went on to develop crash tests that allowed for more detailed understanding of a car's safety features (or lack thereof). This information gave regulators a chance to craft laws incorporating the latest safety findings. Empirical evidence that such laws save lives justified the government getting more involved in proactive risk research. That evidence also made regulators more attentive to the findings of the IIHS and NHTSA. The cycle had started and continues to this day.

In an ideal world, robust research into the risks associated with cars would have coincided (if not preceded) the introduction of the Model T. A rush to innovate, though, won out over safety concerns raised by automobiles.[14] The upshot is that any new risk, theoretically, will always lead to regulation—the timing and efficacy of that regulation, though, depends on whether it is informed by anticipatory, objective, and relevant research or by reactionary pressure to one-off incidents.

---

[10] NHTSA, "Ratings" (accessed Sept. 17, 2023), https://perma.cc/44CZ-57DP; IIHS, "About Our Tests" (accessed Sept. 17, 2023), https://perma.cc/7824-E5VW.

[11] IIHS, "About Our Tests."

[12] See, e.g., NHTSA, "Second Amended Standing General Order 2021-01" (April 2023), https://perma.cc/HM9P-QS4E.

[13] Kevin Frazier, "The Case for Prioritizing the Creation of an AI Benchmarking Consortium," *Lawfare* (Sept. 5, 2023).

[14] Jerry L. Mashaw and David L. Harfst, *The Struggle for Auto Safety* (1990).

What Is Risk Research?

Not all research is equally valuable to the R&R cycle. Pure research is "directed solely toward expanding human knowledge" of a topic.[15] Applied research commercializes foundational research.[16] The line between these two types of research is difficult to spot, which is why scientific endeavors are best thought of as a spectrum.[17]

As research gets closer to the "pure" end, "the degree of uncertainty about the results of specific research projects increases," the goals of that research "become less clearly defined," and outcomes of the research become "less closely tied to the solution of a specific practical problem or the creation of a practical object." Purer research, then, allows for scientists to quickly change the direction of research upon new findings, new technological advances, and new goals. In contrast, research closer to the "applied" end "must be closely constrained by the practical problem which must be solved[.]"[18]

Basic or pure research offers the greatest odds of informing responsive regulation because it has the broadest scope of inquiry—in short, it "extends the knowledge base" of a given technology and its risks.[19] This expansive approach to research increases the odds of discovering risks especially in need of regulation, such as latent risks and unanticipated risks.[20] Research into such risks is generally of minimal commercial value to companies seeking to accelerate the development and deployment of AI and attain profit.[21]

The research most likely to inform AI risk mitigation is closer to the pure research, though not as open ended given its explicit focus on identifying social, political, economic, and environmental consequences of AI proliferation. For that reason, this paper calls for "risk research," research focused specifically on detecting risks from AI. Such research would fill the void in current AI research related

---

[15] Rebecca Eisenberg, "Proprietary Rights and the Norms of Science in Biotechnology Research," 97 *Yale Law Journal* 177, 178 n.1 (1987).

[16] Hans Gersbach et al., "Hierarchical Growth: Basic and Applied Research," 90 *Journal of Economic Dynamics and Control* 434 (2018).

[17] Richard R. Nelson, "The Simple Economics of Basic Scientific Research," 3 *Journal of Reprints for Antitrust Law and Economics* 725, 730 (1971).

[18] *Id.*

[19] Gersbach et al., supra note 16, at 434.

[20] See Camino Kavanagh, "New Tech, New Threats, and New Governance Challenges: An Opportunity to Craft Smarter Responses," Carnegie (Aug. 28, 2019), https://perma.cc/L6GF-KMUF; see generally Chris W. Callaghan, "Surviving a Technological Future: Technological Proliferation and Modes of Discovery," 104 *Futures* 100 (2018), https://perma.cc/8QQC-VDVY.

[21] See Yoshua Bengio et al., "Managing Extreme AI Risks Amid Rapid Progress," *Science* (May 20, 2024).

to "systematic discussion of how to manage long-tail risks from AI systems, including speculative long-term risks."[22]

The Essentials of Risk Research

The quantity, quality, and comprehensiveness of risk research with respect to influencing risk-mitigation policy hinges on several factors. *Quantity* here refers to the number of experiments into the risks raised by an emerging technology. *Quality* pertains to the transferability of research findings into policy. *Comprehensiveness* reflects the extent to which research covers the full geographic and temporal scope of the technology's risks. More experiments of higher quality into the full risk profile presented by an emerging technology require substantial funding, expertise, and independence.

Significant and stable funding is essential for pure research. The absence of any commercial agenda explains why pure research generally relies on public funding.[23] Risk research, too, generally depends on public funds.[24] This type of research simply does not carry the sort of expected returns that private actors look for when setting their budgets.[25] The risks posed by advances in biotechnology, for example, have led researchers to call for more public-sector support for studies of the potential harms of such technology.[26] Without such public support, private-sector research into the capabilities of novel biotechnology, especially how those capabilities may affect less developed countries, has been lacking.[27]

Funding alone, though, will not generate informative and robust research. Risk research also requires relevant expertise in the emerging technology in question. Consider again the IIHS—this small but mighty research outfit includes some of the leading thinkers on auto safety. According to Joe Young, director of media relations at the IIHS, a core group of fewer than ten research engineers and scientists

---

[22] Dan Hendrycks & Mantas Mazeika, "X-Risk Analysis for AI Research," Arxiv (Sept. 2, 2022), https://perma.cc/S7VX-TEE9.

[23] But see Fred L. Smith Jr., "The Basics About Basic Research," 21 *Regulation* 65, 65 (1998).

[24] See Bengio et al., supra note 21.

[25] See "World Leaders Still Need to Wake Up to AI Risks, Say Leading Experts Ahead of AI Safety Summit," Oxford (May 21, 2024), https://perma.cc/HDM3-77Y8 (quoting Jan Brauner, Department of Computer Science, University of Oxford).

[26] Richard Danzig, "Biotech Matters: Public-Private Coordination of Biotechnology," Center for a New American Security (April 22, 2024), perma.cc/Z52M-MJ5Y.

[27] Joel I. Cohen, "Public- and Private-Sector Biotechnology Research and the Role of International Collaboration," in *Managing Agricultural Biotechnology—Addressing Research Program Needs and Policy Implications*, at 134–35, ed. J. I. Cohen (1999), https://perma.cc/4ZCX-FZ3P.

design, implement, and run the tests on new models.[28] These researchers have extensive backgrounds in auto safety and, on average, have worked at the IIHS for decades. By recruiting and retaining such experts, the IIHS has bolstered its reputation and, by extension, the willingness of policymakers to rely on its experiments and recommendations.

A final core component of risk research is independence. If researchers' agendas are shaped by political pressure or financial constraints, they may not achieve the comprehensiveness necessary for robust risk research. The need to appease political supporters and financial backers may explain why life science researchers have generally understudied the risks posed by those emerging technologies to Global South communities.[29] The quality of risk research also may diminish if researchers lack autonomy. Disclosure of a researcher's ties to industry stakeholders may lead regulators and the public writ large to question any findings by that researcher.[30]

The Nature of AI Risk Research Requires an International Organization

The R&R cycle is critical to mitigating risks from emerging technologies. In the AI context, though, it appears as though stakeholders concerned about near- and long-term risks have focused more on regulation than risk research. That imbalance must come to an end. Absent the creation of an international AI risk research initiative that continuously evaluates AI risks, laws governing AI may become outdated or, worse, result in detrimental changes to the technology itself.[31]

AI risks are not easily explained, quantified, nor mitigated. Indeed, some individuals and institutions concerned about AI risks have fragmented into two camps based on their own assessment of the risks most deserving of research and regulation—one camp fears near-term risks such as algorithmic bias and displacement of certain professions; another camp worries about existential risks, or x-risks, such as AI empowering and entrenching totalitarian regimes or enfeebling humans by rendering them overly reliant on AI systems.[32] This debate needlessly distracts from a shared desire to fund research into all risks posed by AI. This research, in turn, can provide a basis for the experts, organizations, and governments

[28] Via emails on file with author.

[29] See generally "Emerging Technologies and Dual-Use Concerns: A Horizon Scan for Global Public Health," World Health Organization (2021), https://perma.cc/6Q7E-HLK5.

[30] See Cristiano Lima-Strong, "A Lawmaker Held an AI Roundtable With Scholars. Most Had Industry Ties," *Washington Post* (May 21, 2024).

[31] "Ossification," in *HTTP/3 Explained* (accessed Sept. 17, 2023), https://perma.cc/F282-TXJB.

[32] See, e.g., Jan Brauner and Alan Chan, "AI Poses Doomsday Risks—But That Doesn't Mean We Shouldn't Talk About Present Harms Too," *Time* (Aug. 10, 2023).

with the capacity to mitigate the most likely and most severe risks identified through rigorous experimentation.

When stakeholders have considered the need for AI risk research, they have prioritized domestic or regional responses over developing an international institution that can provide timely and accurate AI risk analysis for both leading AI development countries and those most susceptible to AI-related harms. That allocation of research and regulatory energy must be reexamined. Subnational or national research entities lack the financial resources, expertise, and independence to meet a global need for AI risk research. In contrast, an international AI risk research initiative may have better odds of meeting these prerequisites than domestic or regional efforts.

Amassing the inputs necessary for cutting-edge AI risk research requires resource pooling. Scarcity plagues three of the most important inputs of AI research—data, compute, and expertise. An international organization is better positioned to resolve those constraints mainly by aggregating financial contributions from its member states; that financial security may also afford researchers within the organization more independence. With more data and compute as well as expert researchers, the resulting risk research could fuel the R&R cycle.

Expensive Inputs—Data and Compute—Require Substantial Funding

AI risk research is quite resource intensive.[33] Specifically, it requires access to training data and compute—two resources that, according to John Etchemendy, co-director of the Stanford Institute for Human-Centered Artificial Intelligence (HAI), are disproportionately held by private actors. Unless and until public actors acquire such resources, Etchemendy fears that "the frontiers of AI [will be left] exclusively in the hands of the most resourced—primarily industry[.]"

On data, labs have used large amounts of data to increase the speed with which their models can learn.[34] It follows that conducting research on AI models would likely benefit from having access to more

[33] "Legislation Will Bolster Public-Sector AI Research," *Stanford Report* (Nov. 6, 2023), https://perma.cc/6QUD-BR88.

[34] See John Etchemendy & Fei-Fei Li, "National Research Cloud: Ensuring the Continuation of American Innovation," Stanford University Institute for Human-Centered Artificial Intelligence (March 28, 2020), https://perma.cc/C5MK-MFX3; Data Foundation, "Charting the Future of Data for AI" (May 16, 2024), https://perma.cc/XNZ5-HTX6.

data[35]—a task made easier by the participation of many actors, especially governments with significant amounts of data that has yet to have been accessed by private AI labs.[36]

On compute, the most advanced labs have leveraged better computational capacity and compute capability to create state-of-the-art AI models.[37] Quality AI research requires similar computational resources.[38] As expressed by researchers at Stanford Law, "the high cost of compute has placed cutting-edge AI research in a position accessible only to key industry players and a handful of elite universities."[39] Graphics processing units or GPUs—one basis to increase compute—are expected to experience yet another price spike.[40] If history repeats, then large AI labs may spend billions on whatever units are produced.[41] Resource pooling, then, is essential with respect to compute. An international research initiative that received funding from governments, private institutions, and a litany of other stakeholders would be better positioned to overcome this cost barrier than any sub-national or national entity—a point made in more detail below.

This reality demonstrates the low odds of any subnational or national research entity keeping pace with the infrastructure spending of big tech companies. Consider that Meta plans to spend billions on compute.[42] No single U.S. state could foot that bill. Even one of the more ambitious federal research

---

[35] Marietje Schaake, "AI and Data Risks: Uniting Voices for a Global Response," UN Trade and Development (April 4, 2024), https://perma.cc/Y9A9-UCH8; see Sethuraman Panchanathan & Arati Prabhakar, "Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource," National Artificial Intelligence Research Resource Task Force at iv (January 2023), https://perma.cc/26E2-J7Z9 (regarding data as key to AI development).

[36] Daniel E. Ho et al., "Building a National AI Research Resource," Stanford HAI, at 9 (2021), https://perma.cc/TLU4-RY8J.

[37] Id.

[38] Schaake, supra note 34.

[39] Ho et al., supra note 36, at 9.

[40] Keumars Afifi-Sabet, "GPU Prices Could Spike Again as Rumors Indicate AMD Wants to Prioritize AI—What Could That Mean for Gamers?" Yahoo (Sept. 17, 2023).

[41] Cf. Kyle Wiggers, "Meta Bets Big on AI With Custom Chips—and a Supercomputer," TechCrunch (May 18, 2023) (flagging that Meta spent billions on Nvidia GPUs but appears to be exploring an alternative means to boost compute).

[42] Jonathan Vanian, "Mark Zuckerberg Indicates Meta Is Spending Billions of Dollars on Nvidia AI Chips," CNBC (Jan. 18, 2024); see Eze Vidra, "The Big Challenge for Generative AI Is GPU Capacity and Server Costs," VC Cafe (April 17, 2023).

proposals is small in comparison;[43] the National Science Foundation will oversee the allocation of $140 million across six different AI research projects, only some of which qualify as risk research.

Even if those financial hurdles were cleared, any domestic AI risk research entity would likely not give equal research access to an international base of experts given the scarcity of data and compute. In turn, countries with insufficient resources to procure such compute will depend on the good graces of wealthier nation-states to do AI risk research on their behalf or, even less likely, to grant their researchers partial access to the nation-state's research infrastructure. Global South stakeholders likely do not want to gamble on such kindness given historical antecedents and the importance of conducting risk research sooner rather than later.

Expertise

On expertise, no country has a monopoly on AI talent.[44] A review of AI-related publications across countries makes clear that expertise resides in several areas. The top-ten list of countries by number of published AI papers from 1997 to 2017 includes China, the U.S., the U.K., Japan, Germany, India, France, Canada, Spain, and South Korea.[45] Other countries have made investments to try to join that group. For instance, Singapore, Australia, New Zealand, and Japan have increasingly strong publication track records and have experienced high rates of growth in the total number of AI professionals in their respective labor markets.[46] An exclusionary or insular approach to AI risk research prevents the inclusion of expertise from different schools of thought, backgrounds, and cultures.

An international approach, beyond permitting broader participation, also has better odds of hiring a sufficient number of experts. While many countries share an interest in expanding their respective AI talent pools, they also face the challenge of directing AI experts to jobs outside of industry. AI labs have successfully lured experts away from the public sector and, especially, academic faculties.[47] The shortage of AI experts on university faculties has significant downstream consequences—the fewer professors capable of teaching AI, the fewer the classes on AI topics, and the fewer the graduates with

---

[43] TRAILS, "New AI Research Funding to Focus on Six Areas" (accessed May 21, 2024), https://perma.cc/D5N9-9SHG.

[44] See Neil Savage, "The Race to the Top Among the World's Leaders in Artificial Intelligence," *Nature* (Dec. 9, 2020).

[45] "China AI Development Report 2018," China Institute for Science and Technology Policy at Tsinghua University, at 12 (July 2018).

[46] Udit Sabharwal et al., "Artificial Intelligence Tech Hubs: Asia Pacific Talent Spotlight" (May 23, 2023), https://perma.cc/5Y48-WV6B.

[47] Brian Eastwood, "Study: Industry Now Dominates AI Research," MIT (May 18, 2023), https://perma.cc/AZ9F-XAWG.

expertise in the field.[48] The aggregation of distributed AI talent at one research entity would not only reduce wasteful competition over experts and reduce the effects of a shortage of AI talent in risk research roles but also ensure that the technical resources made available at the international AI risk research initiative are being used to their maximum potential.[49]

Another barrier to amassing the expertise required to conduct leading AI risk research comes from the broad range of expertise involved in AI research and development. AI projects commonly include "a data scientist, data engineer, machine-learning engineer, product manager, and designer[.]"[50] AI entities operating at the level of the state or even a nation may not have the budget nor personnel from which to recruit the full scope of experts to conduct AI research. California state Sen. Scott Wiener admitted as much upon introducing an AI governance bill—he noted the limited capacity of the California state government to audit AI systems and otherwise implement the proposal.[51] Again, the scale of an international entity can ease the constraints posed by small or tight national labor markets in any and all of these discrete professions.

Independence

An international AI risk research initiative may also be designed to reduce external pressure on its researchers. Comparatively, proposed research institutions with a direct relationship to a subnational or national government often involve a mandate to further innovation, as discussed further below. Commercial concerns and a focus on national competitiveness place a ceiling on a research entity's ability to contribute to a worldwide need for objective, comprehensive, and timely analysis of AI models.

Commercial users of a nation's research entity's resources, for instance, may not want to disclose certain inputs and outputs—same goes for users that have a mandate to prioritize their nation's interests over those of others. This hesitancy among leading AI labs has already been revealed—the Ada Lovelace Institute pointed out that OpenAI withheld certain information about ChatGPT-4 upon its deployment.[52] An expectation of total transparency—at least with respect to the public—goes a step too far, as

---

[48] Remco Zwetsloot & Jack Corrigan, "AI Faculty Shortages," Center for Security and Emerging Technology, at 5 (July 2022) ("[W]e identified a variety of indirect evidence that suggests universities are struggling to meet students' growing demand for AI education.").

[49] Cameron F. Kerry et al., "Strengthening International Cooperation on AI," Brookings (Oct. 25, 2021), https://perma.cc/F5UW-D3GQ.

[50] McKinsey & Company, "New McKinsey Survey Reveals the AI Tech-Talent Landscape" (Jan. 20, 2023), https://perma.cc/9TG3-CDYS.

[51] Billy Perrigo, "California Bill Proposes Regulating AI at State Level," *Time* (Sept. 13, 2023).

[52] Elliott Jones, "Keeping an Eye on AI," Ada Lovelace Institute, at 9 n.6 (July 2023).

recognized by the institute. Perhaps more important than what OpenAI shared is the fact that it exercised complete discretion over that decision. If the research entity relies on corporate investment or participation or has too much of a focus on commercialization, then such omission of information could become a norm—limiting the information available to researchers at the entity as well as other research bodies.

Furthermore, so long as national aims form part of a research entity's mission, the entity may have a harder time soliciting experts from different countries as well as the participation of AI labs based in foreign jurisdictions. For obvious reasons, experts from other countries might think twice before joining or even contributing to any such entity—if they even have the legal authority to do so.[53] Even citizens may refrain from joining such an entity if they thought their work could lead to controversial domestic or national security ambitions.[54]

With the proper governance structure and funding mechanisms, an international AI risk research initiative could exclusively conduct risk research with the participation of experts and AI labs located in a broad range of countries. Moreover, if the initiative manages to develop a "purposive aura" that suggests to researchers, stakeholders, and members of the public that it is meeting an urgent societal need, then the odds of a broad range of participation and engagement by AI experts may increase.[55] Finally, such an initiative—unconcerned with trade secrets and state secrets—could more freely disclose its research methodologies and findings. This level of transparency could help identify any flaws or areas for improvement in the initiative's research.

Why Domestic AI Risk Research Proposals Will Likely Fall Short

A brief review of proposed domestic AI risk research initiatives and of prior efforts by the United States to create an American version of CERN reinforces the case for an international approach to AI risk research.

[53] Executive Order 11935, 41 Federal Register 37301 (1978) (placing limits on federal government hiring of noncitizens).

[54] Emily Badger et al., "The Government Agencies That Became Smaller, and Unhappier, Under Trump," *New York Times* (Oct. 13, 2021).

[55] Arjen Boin et al., "Guardians of Public Value: How Public Organizations Become and Remain Institutions," in *Guardians of Public Value: How Organisations Become and Remain Institutions*, ed. Arjen Boin et al., at 6 (2021).

California

Sen. Wiener proposed a research cloud hosted by the California government that would allow academics and AI startups to access the computing infrastructure necessary "to do advanced AI work."[56] A glance at the initial press release describing the proposal demonstrates the innovation-centric approach to research initiatives sponsored by nation-states or subnational governments. In that release, Wiener explicitly stated that the research cloud—nicknamed CalCompute—would be used to "foster innovation for small businesses[.]"[57] Despite several concerning developments about the safety of AI models since Wiener initially outlined his bill,[58] he continues to frame CalCompute as a "pro-innovation, pro-little-guy" provision.[59] This lines up with Wiener's broader desire to assist with the development of AI—he hopes to "advance the state of the art of an industry that has long called California home," to tap into the technology's "incredible potential to improve people's lives," and to "support [the] massive innovation" that has already taken place in the state.[60]

Even if the bill makes it through the California legislature, it is unclear if the California government would have the resources necessary to fund a research resource capable of doing meaningful risk research. The bill recognizes and anticipates insufficient public resources by authorizing the government to "receive private donations, grants, and local funds[.]"[61] This reliance on external funders could imperil the independence of CalCompute users. Private donors, for instance, could attach strings to their contributions that steer research in a specific direction.

The proposal would lack the potential expertise concentrated within an international initiative. Given that CalCompute would be a project of the California government, it is unlikely that a global set of researchers would be allowed to use the resource, especially because CalCompute is not reserved exclusively for risk research but is also intended for commercial actors. Another limitation of the bill is its timing. The earliest the Wiener proposal could become law is the beginning of 2025.[62]

---

[56] Perrigo, supra note 51.

[57] Scott Wiener, "Senator Wiener Introduces Safety Framework in Artificial Intelligence Legislation," News Release (Sept. 13, 2023), perma.cc/KU9D-RR88.

[58] See Kyle Wiggers, "OpenAI Created a Team to Control 'Superintelligent' AI—Then Let It Wither, Source Says," TechCrunch (May 18, 2024).

[59] Scott Wiener (@Scott_Wiener), X (May 1, 2024, 11:04am), https://perma.cc/W98A-98E8.

[60] Wiener, supra note 57.

[61] SB 1047, Senate Rules Committee Analysis (May 20, 2024), https://perma.cc/K2L6-WZLS.

[62] Perrigo, supra note 51.

NAIRR

Stanford HAI proposed a National AI Research Resource (NAIRR) to facilitate research by academia, government, industry, and civil society users.[63] These users would access "high-end computational resources, large-scale government datasets in a secure cloud environment, and necessary expertise[.]" A congressionally created task force charged with studying the creation of a NAIRR envisions the research entity serving several purposes, including encouraging commercial innovation and advancing the United States's geopolitical interests.[64] The task force's final report, for example, mentions the importance of maintaining "American dominance" in AI research and development.

The proposed NAIRR would rely on the generosity of Congress to sustain its (costly) operations. To the task force's credit, it set forth a thorough budget to ensure as many users as possible would receive access to state-of-the-art AI research infrastructure. The task force envisions an upfront investment of $2.6 billion in a NAIRR to be supplemented annually with a little less than $1 billion.[65] This limited budget may quickly leave users of a NAIRR in a lurch. Even assuming the task force received its full budget request (a big assumption),[66] the resulting NAIRR would serve a relatively small user community—about 19,000 users working on around 2,300 projects.[67]

The combination of NAIRR leaders needing to constantly justify to Congress continued massive appropriations—to buy GPUs, retain staff, and the like—and the NAIRR already having an explicit innovation mandate suggests that researchers will lack full independence over their research agendas. The drafters of the resource tout its "potential not only to unleash a string of advancements in AI, but to help ensure the U.S. maintains its leadership and competitiveness on the global stage." Congress will expect researchers to live up to that potential, even if it means focusing less on risk research.

As a U.S. research entity, the NAIRR would also have a smaller talent pool in comparison to an international organization. The resource would operate pursuant to U.S. interests and exclusively serve users based in the U.S. or affiliated with U.S. organizations. Moreover, governance decisions around the use of a NAIRR would incorporate "interests and perspectives from across Federal agencies."

---

[63] Stanford HAI, "Stanford University Human-Centered AI Proposed National AI Research Resource" (accessed Sept. 17, 2023), https://perma.cc/C7TU-VK29.

[64] NAIRR Task Force, "Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem," at v (2023), https://perma.cc/G9DV-ZYC9.

[65] Id., at 49.

[66] See, e.g., Alan Rappeport, "U.S. National Debt Tops $33 Trillion for First Time," *New York Times* (Sept. 18, 2023).

[67] NAIRR Task Force, supra note 64, at 48.

A NAIRR would also take several years to get off the ground. So even if issues related to its funding, expertise, and independence were resolved, it would still be quite some time before the resource could produce research to inform the R&R cycle.

Neither CalCompute nor a NAIRR is a "bad" idea. This paper argues instead that those efforts are insufficient with respect to generating the research—and, by extension, regulation—necessary to mitigate AI risks wherever they inevitably manifest.

The Limits of Sustained Federal Investment in Pure or Risk Research

Advocates for CalCompute or NAIRR should also heed the lesson of the Brookhaven National Laboratory (BNL). A brief summary of the rise and fall of efforts to expand the BNL's research portfolio reinforces the limitations associated with a subnational or national research initiative.

An examination of the United States's struggles to undertake world-leading particle physics at the BNL also bolsters the case for studying CERN as a template for international research coordination. The importance of scale—with respect to funds and expertise—and a collaborative culture to CERN's work helps to explain why similar, albeit more insular particle physics labs never got past the ideation phase in the United States.

The U.S. tried in the early 1980s to move ahead with updates to the BNL, which conducts fundamental particle physics research similar to that conducted at CERN. Efforts to scale up the BNL, however, ran out of steam while climbing Capitol Hill. The *New York Times*, perhaps providing a lens into the perspective of a majority of Americans, pointed out that upgrading BNL would require (a) too much money upfront, (b) too much ongoing financial investment, (c) too much time to develop if the U.S. aimed to keep pace with CERN, and (d) too much political capital, which BNL scientists could likely not generate.[68] The *Times* called for the U.S. to instead double down on its participation in CERN.

A lack of political will stifled similar proposals throughout the 1980s. Moreover, physics no longer enjoyed supremacy as the field most worthy of scientific funding. Shinier objects came to the fore. The U.S. federal government increasingly struggled to decide "which sciences should be awarded the resources to delve into ever more arcane—yet ever more fundamental—areas at a time when the cost of research [was] soaring and the economy [was] not."[69]

By the late 1980s, looking back at decades of failed attempts to get the U.S. on the level of CERN, the *Times* argued that "America has no coherent policy for supporting scientific research." By way of example, the *Times* examined Congress's consideration of a proposed "superconducting super collider," or SSC, a $5 billion machine for particle physics experiments. Congress, per the *Times*, had not assessed

[68] "The Trouble With Isabelle," *New York Times* (June 24, 1982).

[69] Robert Crease & Charles Mann, "Gambling With the Future of Physics," *New York Times* (Dec. 5, 1982).

the trade-offs that the SSC would require. The proposal seemed to ignore other obvious options to achieve similar research—namely, greater participation in CERN. In short, the *Times* chalked up the SSC to an idea that may have been justified in "different times."

When the 1990s rolled in, the dearth of political will for "big science" became obvious. In 1992, the U.S. House voted to halt funding for the SSC.[70] Representatives likely took issue with the fact that CERN was working on a similar collider at nearly one-eighth the cost of the SSC. Those opposed to the SSC also noted that pure research may not deserve investment when evaluated against projects likely to "ease pressing problems of the world." CERN researchers speculated that U.S. legislators may have recognized that Europe had successfully executed a trans-Atlantic brain drain that meant more of the United States's top physicists did research in Europe than vice versa. Reversing that trend, of course, would add to the costs and logistics of getting the SSC going.

This brief examination of the United States's struggles to go it alone on "big" science reinforces that such endeavors require scale in more ways than one. In particular, the United States's shortcomings demonstrate that a large gross domestic product and a place atop the global political order cannot guarantee the success of such initiatives.

* * *

Ongoing, high-quality, independent research into the global and intergenerational risks posed by AI is unlikely to occur within a domestic research organization. Even a risk research resource hosted by the U.S. government, such as the proposed NAIRR, would lack the financial resources, expertise, and independence to produce risk research of a similar quantity, quality, and scope as an international AI risk research initiative.

Not just any international entity, though, would realize those outcomes. Many international organizations have earned less than stellar reputations.[71] Study of CERN and the IPCC, two international research-focused bodies that have earned acclaim as model organizations, provides important insights for how an international AI risk research initiative could be designed to realize the unique advantages of a global organization.

[70] Barry James, "Europe Is Ready to Pick Up the Pieces in Particle Research," *New York Times* (June 20, 1992).

[71] See generally Hylke Dijkstra & Maria J. Debre, "The Death of Major International Organizations: When Institutional Stickiness Is Not Enough," 2 *Global Studies Quarterly* 1 (2022).

An International AI Risk Research Initiative Should Learn From CERN and the IPCC

Many advocates for more AI risk research have proposed a "CERN for AI" or "IPCC for AI" in recognition of the exemplary work of those organizations.[72] Yet few have explored in detail what makes CERN and the IPCC successful research bodies and to what extent those causes can and should be copied in the AI context.

CERN and the IPCC stand out as models for the development of an AI risk research entity because each organization earned recognition as an "institution" that produces valuable scientific insights. A CERN for AI would entail the creation of a centralized hub for AI research, and the aggregation of resources and expertise by nations around the globe would make cutting-edge AI research possible. This worthy goal, though, faces numerous barriers that render it unlikely that a CERN model could be exactly replicated in the AI context. An IPCC for AI, however, is a more realistic model that could still result in tangible reductions in AI risk. This approach would involve concentrated sprints in which a global set of AI experts analyzed the latest AI research and, with guidance from national governments, produced consensus reports on the most pressing AI risks and potential steps to reduce those risks.

This part introduces the distinguishing aspects of each institution, identifies structural and cultural features worthy of incorporation into an international AI risk research initiative, and pinpoints barriers to such an initiative replicating either institution. The takeaway is that an international AI risk research initiative should and *must* be something novel, while paying respects to the lessons provided by CERN and the IPCC.

CERN

CERN operates several cutting-edge particle accelerator facilities with the goal of "uncover[ing] what the universe is made of and how it works."[73] The organization emerged from a unique historical moment that facilitated robust national cooperation and encouraged substantial investment in "big" scientific endeavors.[74] Decades later, it serves as a model for how to break through the gridlock that, with

[72] David Matthews & Martin Greenacre, "EU Science Advisers Back Call for a 'CERN for AI' to Aid Research," Science Business (April 15, 2024); Holger Hoos & Morten Irgens, "Europe Needs a CERN for Artificial Intelligence," Science Business (Oct. 24, 2023); Mustafa Suleyman et al., "Proposal for an International Panel on Artificial Intelligence (AI) Safety (IPAIS): Summary," Carnegie (Oct. 27, 2023), https://perma.cc/N6RK-7ZPK.

[73] CERN, "Our Mission" (accessed Sept. 17, 2023), https://perma.cc/XZR4-KCSR.

[74] Jos Engelen & Paul 't Hart, "CERN: Guardian of Human Aspiration to Understand the Universe," in *Guardians of Public Value: How Public Organisations Become and Remain Institutions*, ed. Arjen Boin et al., at 214–18 (2021); Article II, § 1, CERN Convention, https://perma.cc/JQD5-LSRN.

increasing frequency, undermines international collaboration.[75] For the purposes of this paper, CERN embodies the possibility of nation-states jointly pursuing pure research, ceding (for the most part) control over that research to scientists, and committing to a flexible, yet sustainable funding mechanism that permits the organization to serially invest in cutting-edge infrastructure to maintain its reputation as the foremost hub of particle physics research.

Several attributes of CERN have contributed to its success. In particular, its structural protections of collaboration, funding, and independence empower the organization to generate useful research. This section provides an overview of those attributes as well as an assessment of whether a contemporary international AI risk research initiative could emulate those attributes.

Cohesion, Collaborative, and Clarity of Mission

CERN launched in the wake of World War II when the will to form international governance institutions likely reached an all-time high.[76] That said, CERN did not initially include a broad range of actors. Even at a time of heightened willingness among nations to form collaborative, international bodies, just eleven European nations can call themselves CERN founding members.[77] What's more, those eleven were relatively geographically proximate, held similar aspirations, and shared certain values and historical ties.

The unique cultural and geopolitical conditions that existed at the birth of CERN hinder its relevance to shaping a contemporary international research organization. Today, geopolitical tensions have resulted in "the inability of countries to cooperate via international institutions to address policy problems that span borders."[78] Beyond the difficulties of forming an international entity posed by a competitive and hostile international climate, the initiative would face other barriers that did not apply to CERN. The initiative likely would need to involve more (likely many more) countries as well as other types of actors—such as AI labs.[79] And, whereas CERN participants shared an understanding of the purpose of the organization and its potential outcomes, AI research brings up more questions than answers—and those questions involve several politically and economically significant topics that, once answered, could change humankind.

---

[75] Mark Robinson, "The CERN Community: A Mechanism for Effective Global Collaboration?" 10 *Global Policy* 1, *passim* (2019).

[76] See CERN, "The History of CERN" (Sept. 17, 2023), https://perma.cc/XUN7-TCCT.

[77] CERN, "Our Member States" (accessed May 21, 2024), https://perma.cc/SRG5-ZPTE.

[78] Robinson, supra note 75, at 1.

[79] Gary Marcus & Anka Reuel, "The World Needs an International Agency for Artificial Intelligence, Say Two AI Experts," *Economist* (April 18, 2023).

The impediments to an initiative having cohesive and collaborative membership cannot be bypassed. Cohesion among CERN member states facilitated many of the key structural decisions that have contributed to its decades of world-leading science. Once the group decided to move forward with CERN, they did so with haste. The member states quickly agreed to a convention that afforded researchers tremendous discretion and provided them with sufficient resources, limited the influence of national governments, and prevented the possibility of mission creep into militaristic or commercial affairs.[80]A quick dive into the core elements of the convention gives a sense of the "bones" that have upheld CERN through economic and political turmoil.

The who, what, and why of CERN's research agenda are explicitly outlined in the CERN Convention. Article II, Section 1, of the convention states that "[t]he Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character[.]" A couple things about that mission deserve attention: First, collaboration is mandated; second, the geographic scope of participants is limited; third, the subject and purpose of research are made clear; and, fourth, a different part of the article unequivocally prevents CERN from having "concern with work for military requirements[.]"

A mandate for collaboration also appears in another part of the convention. Section 7 of Article II directs CERN to "cooperate to the fullest possible extent with laboratories and institutes in the territories of member states within the scope of their programmes and activities."[81] This direction evidenced a desire to "avoid duplicating research work[.]" The explicit instruction to avoid wasteful research resulted in CERN becoming the focal point of a broad global network of researchers—thereby making CERN the indispensable center of international physics exploration.

With respect to an international AI risk research initiative, emulating these core parts of the convention seems unlikely. Consider the prohibition on commercial and military aims—CERN's pure scientific focus has fostered collaboration among researchers and shielded the organization from quarrels over things like revenue-sharing agreements. On paper, it is easy to espouse a comparably simple mission for the international AI initiative such as the identification of risks and analysis of means to mitigate those risks. In practice, the willingness of nations and other potential participants to meaningfully contribute expertise and resources to such a narrow mission has yet to be determined. Additionally, how best to fulfill even that narrow mission is not clear.

Unlike debates over the size and scale of new particle colliders, AI research could take several and significantly different forms. For instance, would AI risk identification solely involve evaluating models submitted by AI labs? Or would development of means to reduce those risks be within the initiative's bailiwick as well? Anka Reuel, a doctoral student at the Stanford Intelligent Systems Laboratory at Stanford University, and Gary Marcus, emeritus professor of psychology and neural science at New

[80] See, e.g., Article II, § 1, CERN Convention.

[81] Article II, § 7, CERN Convention.

York University, for instance, make the case for the development of "automated or semi-automated tools for answering fundamental questions, such as 'How much misinformation is out there?', 'How rapidly is its volume growing?' and 'How much is AI contributing to such problems?'" It is possible that an international AI risk research initiative would have the requisite resources and expertise to develop those tools, but with every expansion of the initiative's mission there may be a trade-off in the willingness of members to fully participate in that mission.

With that hesitancy in mind, the CERN Convention's mandate of "cooperat[ion] to the fullest extent possible" may also cause friction if made a nonnegotiable condition to an actor's participation in an international AI risk research initiative. Whereas the line from particle collisions to weaponization of that research is attenuated (if not nonexistent), national security observers have made clear that AI advances can and, according to some, must be integrated into a nation's defense strategy.[82] The same is true with respect to commercialization. Some countries, like the United Kingdom, have signaled a "pro-innovation" stance on AI development.[83] Such nations may not voluntarily cede data, expertise, and other inputs that could assist with the initiative's aims.

Calls for a CERN for AI must take this complexity head on. An unambiguous and simple AI research agenda combined with structural limitations on mission creep may be harder to identify than in the CERN context. But, once identified, that mission can serve as a source of inspiration for experts, nation-states, AI labs, and other stakeholders to work together on the creation of an international AI risk research initiative.

Stable and Significant Funding Plus Resource Consolidation

CERN member states, from the outset of the organization's creation, not only complied with the convention but also made several significant voluntary actions to shore up CERN's legitimacy and capacity. In particular, they committed not only to continue to pool their resources but also to automatically increase their annual financial contributions to CERN's budget.[84]

The financial security of CERN also received a boost from members' agreement to an innovative funding mechanism: The required contribution by each member depends on (a) the programs they participate in (i.e., the research they conduct and support) and (b) the size of their economy—specifically the "average net national income at factor cost of each member state for the three latest

[82] See, e.g., Chuck Young, "How Artificial Intelligence Is Transforming National Security," U.S. Government Accountability Office (April 19, 2022), https://perma.cc/PYA7-VN6A.

[83] See, e.g., Astha Rajvanshi, "Rishi Sunak Wants the U.K. to Be a Key Player in Global AI Regulation," *Time* (June 14, 2023).

[84] Dominique Pestre, "Some Characteristic Features of CERN in the 1950s and 1960s," in *Studies in CERN History*, 1st. ed., at 3 (1989), https://perma.cc/ZWS8-XG6L.

preceding years[.]"[85] This mechanism anticipates and tolerates member states occasionally falling behind on their contributions. Flexible payment plans allow such states to remain CERN stakeholders, facilitating the continued participation of a growing set of member states.[86] Members, though, have not exploited this flexibility—in fact, some states derive their annual CERN dues from the foreign policy budget as a way to safeguard their contribution from political whims given that such budgets are typically less likely to experience significant cuts.[87] This collective willingness to bolster CERN's prospects also manifests in members' acceptance that their contributions come with few explicit benefits. For instance, there is no guarantee of a "fair return" or that the nation's share of CERN expenditures will be reinvested in that nation's economy.[88]

On the whole, this approach to funding has been a major part of CERN's evolution into an "institution." An alternative approach would likely have saddled CERN's work "because of the enormous difficulty of foreseeing one's needs in a field where the leading edge of research evolved extremely rapidly."[89]

Another pivotal step in CERN's development and entrenchment as the "apex" of the European scientific community came when several members mothballed their respective national physics labs and directed those resources and personnel to CERN instead.[90] The result was CERN having the financial and political capital to recruit and retain experts across the continent and, later, the world. In other words, members consolidated their respective particle physics resources into a central location. This consolidation has been instrumental to CERN's research. Around 12,000 researchers from more than 70 countries conduct work together to develop novel experiments that generate data analyzed by 170 data centers spread across 36 countries.[91] This distributed yet CERN-specific research community helps the organization retain a "very competent staff," "offer attractive employment conditions," and accomplish "[a] series of landmark achievements demonstrated by the added value of the collaboration across disciplinary and national boundaries."

A similar level of resource consolidation seems unlikely in the AI context given the national security imperatives and economic significance national leaders have placed on achieving AI superiority. South

[85] Article VII, CERN Convention.

[86] See, e.g., Edwin Cartlidge, "Greece May Not Be Able to Afford CERN," *Science* (April 3, 2012) (discussing flexible payment plans for Greece to remain affiliated with CERN through its austerity crisis).

[87] See Robinson, supra note 75, at 3.

[88] See Pestre, supra note 84, at 4.

[89] Id., at 3.

[90] See Pestre, supra note 84, at 6.

[91] Engelen & 't Hart, supra note 74, at 220.

Korean leaders intend their country to become one of the three leading AI "powerhouses" by 2027.[92] President Emmanuel Macron of France has similar ambitions for his country. He recently pledged $500 million to spur the development of AI "champions."[93] President Biden has likewise stressed that he wants the U.S. to "lead the way toward responsible innovation."[94] That goal may foreclose substantive resource consolidation.[95] The Biden administration has taken measures to deny China access to GPUs in order to advance the United States's standing in the competition for "supremacy" in artificial intelligence. U.S. allies have even been drawn into this resource skirmish. Pressure from the U.S. resulted in the Netherlands and Japan paring back the export of certain AI tools to China. Rather than stymie China, though, the U.S. may have only accelerated Chinese investment in AI inputs and given it cause to hold back from any sort of resource sharing with respect to international AI research.

A willingness to share resources and expertise does not exist in the present AI climate. Even among "friendly" nations, maintenance of competitive barriers rather than the facilitation of collaboration is the norm. For sake of illustration, in an analysis conducted by the Carnegie Endowment for International Peace of how best to foster more science and technology collaboration between the U.S. and Japan, several barriers—each intended to protect the interests of the respective nations—became clear.[96] Case in point, U.S. law prohibits Japanese nationals from contributing to U.S.-sponsored sensitive scientific research. Even if those security clearance issues were resolved, additional collaborative tasks—such as the exchange of sensitive information—would likely entail numerous and significant changes to Japanese law to comport with U.S. standards and to alter U.S. perceptions about the inadequacy of Japan's information security ecosystem.

Scarcity of AI research resources mandates resource pooling just as the logistical and financial challenges of building and continuously operating a particle collider required nations to coordinate and sacrifice certain national assets. For the reasons spelled out above, nations likely will hold more resources back from an international AI risk research initiative than the founding members of CERN when launching that effort. To achieve financial stability and to become a hub of AI expertise, a CERN for AI will have to broaden the sources from which it pools resources—a group of eleven states will not do. Attempting to coordinate resource pooling from more actors as well as from more diverse actors will

---

[92] Sheila Chiang, "South Korea Wants to Be a Top A.I. Hub—Its Memory Chip Dominance Could Be an Advantage," CNBC (July 5, 2023).

[93] "Macron Announces €500 Million in Funding for AI," *Le Monde* (June 14, 2023).

[94] Molly Nagle, "Biden White House, Tech Companies Launch New Safeguards Around Emerging AI Technology," ABC (July 21, 2023).

[95] Dave Lawler, "How the U.S. Is Trying to Stay Ahead of China in the AI Race," *Axios* (June 29, 2023).

[96] James L. Schoff et al., "A High-Tech Alliance: Challenges and Opportunities for U.S.-Japan Science and Technology Collaboration," Carnegie Endowment for International Peace (July 29, 2021), https://perma.cc/NAW2-AXS9.

pose challenges. Yet, given the importance of such research being global, being well endowed, and being highly transparent, no alternative will do.

Separation of Science and State: The Ideal Governance Model for Independent Research

Governance at CERN also developed in a way that maintained and furthered the goals of the convention and those of the member states more broadly. In brief, CERN relies on a flat governance structure led by members of the scientific community. The CERN Council oversees the organization and has the authority to make critical decisions[97]—it "controls CERN's activities in all matters, scientific, technical and administrative."[98] Two representatives from each of the member states—one science administrator and one leading scientist—make up the council. A council president is selected by the council to serve a single three-year term. This relatively short tenure ensures that member states have (or at least perceive that they have) a fair amount of control over CERN's direction.

Each member state delegation has one vote in council decisions. This voting practice diverges from similar international organizations, such as the International Monetary Fund, that allocate voting power in proportion to financial contributions.[99] Scholars contend that the "one member, one vote" approach has created a "stable, level-playing field" among member states.[100] The retention of this voting system through the expansion of the number of member states from twelve to twenty-three—many of which contribute vastly different amounts in terms of financial resources and technical expertise—provides evidence for that theory.

A preference for achieving consensus among council members may also explain the longevity of the voting system. Though the convention permits the council to make nonunanimous decisions, the council aims for consensus—and usually achieves it. In other words, a norm of unity serves as a carrot for everyone to find common ground, and the rules act like a stick that reminds members that dissent may provide little value other than sowing unproductive discord.

CERN's governance structure also includes a director-general (DG) who acts as the organization's CEO and legal representative.[101] Whereas the council tends to advance the intergovernmental and policy-

[97] CERN, "Our Governance" (accessed Sept. 20, 2023), https://perma.cc/4UAU-P5YQ.

[98] Engelen & 't Hart, supra note 74, at 221.

[99] International Monetary Fund, "IMF Members' Quotas and Voting Power, and IMF Board of Governors" (accessed Sept. 21, 2023), https://perma.cc/DUN4-C68X.

[100] Engelen & 't Hart, supra note 74, at 221–22.

[101] CERN, "Welcome to the CERN Council" (Sept. 21, 2023), https://perma.cc/Z2SW-QDME.

driven aspects of CERN, the DG—informed by "elite scientists whispering in [their] ear"—complements the council by focusing on supranational and scientific priorities.[102]

The council appoints the DG—usually for a single five-year term.[103] This appointment process typically involves the selection of one of several candidates with very similar profiles: "European physicists who [have] already made their mark [in the field] and who [have] already had some experience with the intricacies of CERN management."[104] To ease the transition from one DG to the next, the incoming DG is appointed a year prior to expiration of the current DG's term.[105] The combination of a single term and a built-in transition period diverged from the status quo and may have played a role in CERN's experiencing relatively more stability than similarly oriented organizations.

Several advisory bodies play a role in CERN governance, including the Scientific Policy Committee and the Finance Committee. Committee members are elected by their peers exclusively on merit and entirely independent of their nationality and other affiliations. These committees serve an advisory function and often as incubators for the next generation of CERN leaders. CERN also hosts several "user committees" that develop as needed around the operation of specific machines and send representatives to the Advisory Committee of CERN Users, which more generally advocates on behalf of CERN users.[106] All of these committees lack formal authority, but they meaningfully contribute to CERN's "strong corporate spirit" as well as to CERN's governance bodies having "inimitable capabilities to adapt to emerging issues and shifting constellations of power and interests."[107]

On the whole, CERN has become a "nimble, cooperative, science-driven organization" due to its flat and simple governance model and its mandate to adjust its operations in pursuit of a clear, specific, and unchanging mission.[108]

Replication or even partial emulation of such a governance structure by a CERN for AI is a tall task that depends on the extent to which members can agree on a specific research mission and their willingness to cede control over their pooled resources to experts. As mentioned throughout this paper, realizing those conditions will require extraordinary acts of leadership and collaboration. With a risk research

[102] Engelen & 't Hart, supra note 74, at 217.

[103] Cf. CERN, "CERN Council Appoints Fabiola Gianotti for Second Term of Office as CERN Director General," Press Release (Nov. 6, 2019), https://perma.cc/A35F-BB6D.

[104] Pestre, supra note 84, at 3.

[105] Robinson, supra note 75, at 8.

[106] CERN, "ACCU: A Committee That Addresses Users' Needs" (June 4, 2013), https://perma.cc/ZN6E-CB9U.

[107] Engelen & 't Hart, supra note 74, at 220.

[108] Id., at 230–31.

mission in place and the identification of specific experiments to realize that mission, however, it may become easier for governments, labs, and other AI stakeholders to afford the international AI risk research initiative more autonomy. In that scenario, nations may be more willing to share data, compute, and expertise because they have assurance that those resources will not directly benefit the commercial and militaristic aims of any other member.

If that two-step process occurs, the next step is to develop a governance structure that, like CERN's, protects experts from the shifting mandates and priorities of members. CERN's record of success lends support to the "one member, one vote" model as a means to dampen political winds and to foster a sense of collective decision-making among members. This latter outcome may be particularly important if the governing body includes nations that, though currently lacking substantial financial resources and technical expertise, stand to bear a disproportionate amount of the losses generated by mismanagement of this common-pool hazard. By way of example, analysts expect that disinformation created and spread faster by AI will have more disruptive effects in the Global South than the Global North.[109]

IPCC

The IPCC writes "authoritative reports [that] inform international policy and negotiations on climate change."[110] Since it published its First Assessment Report in 1990, the organization has "developed into a unique global intergovernmental body, with the hybrid quality of being both scientific and political." Particularly relevant here is the IPCC's creation of a global network of experts and implementation of processes that allow member states to achieve consensus on complex and controversial scientific matters.

A few characteristics of the IPCC deserve further study by those seeking to develop an international AI risk research initiative. This section covers those characteristics and examines their applicability to a new AI entity.

Consensus Can Drive Action

Rather than operating as a centralized hub that generates research, the IPCC serves as a consolidator, verifier, and summarizer of research. In the (likely) event that potential members of an international AI risk research initiative are not willing to lend their financial resources and technical expertise to an AI research hub akin to CERN, then the IPCC's model likely warrants especially close attention by

[109] Carolina de Assis, "Global South Faces 'More Pronounced Challenges' Integrating Artificial Intelligence With Journalism, New JournalismAI Report Finds," *LatAm Journalism Review* (Sept. 27, 2023), https://perma.cc/67BE-F7V8.

[110] Eric Paglia & Charles Parker, "The Intergovernmental Panel on Climate Change: Guardian of Climate Science," in *Guardians of Public Value: How Organisations Become and Remain Institutions*, ed. Arjen Boin et al., at 295 (2021).

advocates for an international AI risk research initiative. This alternative would not produce the same agglomeration effects as having the world's foremost AI experts spend time together, collaborate on specific projects, and challenge one another's perspectives. Still, action by the international community in response to the IPCC's assessments makes clear that the IPCC's model of establishing consensus about the sources, magnitude, and severity of global risks can facilitate meaningful regulation of those risks.

The IPCC methodically drafts reports that provide "a comprehensive, objective and transparent assessment of the current state of knowledge of the science related to climate change."[111] The drafting stage involves the formation of "author teams" tasked with producing a topic-specific portion of the report.

The author selection process proceeds in two steps. First, governments and IPCC observer organizations nominate authors—this includes the submission of detailed CVs.[112] Second, the IPCC selects the authors based on their expertise and in consideration of the organization's goal of ensuring the participation of experts from a range of specialties, backgrounds, and values. As a result, author teams include a mix of men and women, young scientists and more senior researchers, and individuals with industry experience as well as individuals from nonprofit organizations.

The selected authors then produce a first report following their review of "scientific, technical and socio-economic literature in scientific journals and other relevant publications."[113] Through the drafting process, the authors try to produce a "[b]alanced assessment of the full range of scientific views" as well as to shield their product from "the influence of special interests[.]" The IPCC maintains that their "method of author team selection, multiple rounds of review of each report, and ... Conflict of Interest Policy" make that balance possible.[114]

The first drafts then undergo review by hundreds of experts. By way of example, 659 experts reviewed Working Group I's first draft for their contribution to the Fifth Assessment Report.[115] Reliance on experts, though, does not mean this process is not open and participatory. The organization has designed a consultation system that "faciliat[es] the participation of [as many] experts encompassing as wide a range of views, expertise and geographical representation as possible." This includes scholars, industry

---

[111] IPCC, "IPCC Factsheet: How Does the IPCC Review Process Work?" at 1 (2015), https://perma.cc/K35D-4P9N [hereinafter, IPCC Factsheet].

[112] IPCC, "How Does the IPCC Select Its Authors?" at 1 (accessed May 21, 2024), https://perma.cc/9SZS-CABA [hereinafter, IPCC Author Selection].

[113] IPCC Factsheet, supra note 111, at 1.

[114] IPCC Author Selection, supra note 112, at 2.

[115] Id.

representatives, nonprofit experts, and others.[116] Individuals intent on commenting can simply go through a self-declaration process to establish their credentials as an expert.[117]

IPCC's deliberate and extensive outreach also characterizes the remainder of the drafting process. The IPCC explains that "[a]fter the expert review of the First Order Draft, author teams prepare a Second Order Draft of the report, taking into account the review comments received; a first draft of the report's Summary for Policymakers (SPM) is also prepared."[118] Clearly, the initial round of review comments are not for show nor are the preliminary stages intended only for highly technical audiences. Once the Second Order Drafts and SPM are ready, the IPCC invites the first round experts to comment once more, invites other experts to do the same, and also provides governments with a chance to review and comment on the drafts. These, again, are not empty gestures. Manifold stakeholders take advantage of these input opportunities—illustrated again by engagement with Working Group I's section of the assessment: Exactly 800 experts commented on the Second Order Draft; twenty-six governments did the same.

This consultation process includes numerous periods of outreach and several opportunities for comment by experts and governments. The thoroughness of that outreach and importance of the IPCC's work to those stakeholders is highlighted by the more than 142,000 comments that were submitted during the drafting of the IPCC's Fifth Assessment Report.

All of this work is at the direction and mandate of the IPCC Panel—a body of 195 member governments.[119] The panel operates on a consensus basis to "decide on the organization's budget and work program; the scope and outline of its reports; issues related to principles and procedures of the IPCC; and the structure and mandate of the IPCC Working Groups[.]"[120]

By permitting nations that would benefit from a less dire assessment to poke and prod at its findings, the IPCC has established a reputation for "[s]cientific integrity[.]" The value of that integrity cannot be understated—Eric Paglia, a researcher at KTH Royal Institute of Technology, and Charles Parker, a professor at Uppsala University, opine that the IPCC's integrity is "the basis of the [IPCC's] legitimacy and epistemic authority as well as the source of its policy impact."[121]

---

[116] Nick Stockton, "The US Won't Pay for the World's Best Climate Science," *Wired* (Aug. 11, 2017).

[117] IPCC Author Selection, supra note 111, at 1.

[118] Id.

[119] IPCC, "Structure of the IPCC" (accessed Sept. 21, 2023), https://perma.cc/XR6D-ZYDL

[120] Id.

[121] Paglia & Parker, supra note 110, at 296.

Absent a consensus-based approach, concerns about the IPCC struggles with overrepresentation of some views might completely undermine its effectiveness, rather than slightly hinder it. More specifically, the IPCC's thorough and participatory drafting process reduces the validity and spread of concerns held by some observers that the IPCC lends too little weight to climate research produced by the social sciences and humanities or that IPCC assessments "[o]ver-represent[] ... knowledge produced in industrialized countries[.]"[122]

The proposed international AI risk research initiative must embrace the disparate values, perspectives, and findings regarding the likelihood and severity of different AI risks. By way of example, scholars most concerned about short-term risks as well as those most concerned about long-term risks should have their work reviewed by the initiative and be engaged in the production of its reports.[123] If the initiative—by personnel, process, or both—appears predisposed to conclude that certain types of risks deserve more attention than others, then certain experts may refrain from participating in the initiative and AI stakeholders may hold back from acting on the initiative's conclusions. Of course, achieving consensus often comes at a cost. The IPCC's record of success, however, suggests that "adversarial scrutiny" of contested and controversial topics makes up for qualifications of findings that might have otherwise been stronger.

The Limits of Volunteerism

The IPCC's funding mechanism and expert selection processes structure tell a cautionary rather than an exemplary tale. About twenty-five countries voluntarily fund the IPCC's "shoestring operations."[124] The majority of the IPCC's budget goes toward covering flights for experts to attend one or two working group meetings in advance of that assessment.

This limited donor base and budget means that any change in support can have drastic effects on an institution with limited financial wiggle room. Case in point, in 2017, the Trump administration eliminated U.S. support—depriving the IPCC of $2 million or nearly half of the IPCC's annual operating expenses.[125] Though the IPCC managed to circumvent the resulting funding shortfall, it did so by placing a greater financial burden on other countries and by tapping into limited financial reserves.[126] If the international AI risk research initiative is going to reliably produce consensus reports on the risks

---

[122] Id., at 297.

[123] Kif Leswing, "Parrots, Paper Clips and Safety vs. Ethics: Why the Artificial Intelligence Debate Sounds Like a Foreign Language," CNBC (May 21, 2023).

[124] Stockton, supra note 116.

[125] Id.

[126] Brenda Ekwurzel, "Donald Trump Ends IPCC Funding and 'Abandons Global Science Leadership,'" *Ecologist* (Aug. 17, 2017).

posed by the latest AI models, then it will need to make investments in recruiting and retaining AI experts as well as in the technical infrastructure necessary to analyze the latest AI research.

The need of an international AI risk research initiative for more reliable and significant funding is the sum of at least two factors. First, the scarcity of AI experts may militate against a general willingness to volunteer for the initiative. AI experts have no shortage of opportunities, many of them lucrative, to share their two cents. The initiative would likely not be able to count on the good graces of those experts to volunteer weeks or months of their time. Second, verification and summarization of AI research may require more capital investments than the same task when done in the climate science context.

The AI initiative must also develop a different approach to expert selection. The IPCC's reliance on voluntary participation by experts may be possible only because of the time between each of its assessments—typically six to seven years.[127] A much shorter period might cause fewer experts to lend their time. Interviews of IPCC experts confirm that they find the work "intense, stressful, and unsustainable."[128] Yet an international AI risk research initiative likely has no other choice than to produce consensus reports on a much shorter timeline if the results are going to inform regulatory efforts.[129]

The IPCC relied on an unrepresentative set of experts to produce its first assessment[130]—nevertheless, the international community accepted and acted on the report as evidenced by the report having "played a pivotal role in the creation of the [United Nations Framework Convention on Climate Change], the centerpiece of the global climate change policy regime."[131] Of the one hundred or so authors who contributed to that first assessment, fewer than twenty authors represented the Global South.[132]

An AI research report issued by a similar group of authors would likely receive a far different reception today. It is unclear why, despite the unrepresentativeness of the First Assessment Report authors, the report carried as much sway as it did. One theory is that because the international community created the IPCC and oversaw the issuance of the first report, those diversity concerns were not of paramount

---

[127] IPCC, "IPCC Meets to Approve the Final Component of the Sixth Assessment Report," Press Release (March 13, 2023), https://perma.cc/P2S6-ULLZ.

[128] Ayesha Tandon, "Analysis: How the Diversity of the IPCC Authors Has Changed Over Three Decades," Carbon Brief (March 15, 2023).

[129] See, e.g., Will Henshall, "4 Charts That Show Why AI Progress Is Unlikely to Slow Down," *Time* (Aug. 2, 2023); Shana Lynch, "AI Benchmarks Hit Saturation," Stanford HAI (April 3, 2023), https://perma.cc/AR2Q-S6U7.

[130] Tandon, supra note 128.

[131] Paglia & Parker, supra note 110, at 296.

[132] Tandon, supra note 128.

importance. Regardless of the rationale, the international AI risk research initiative likely cannot and should not assume similar treatment should it fail to, from the outset, include a globally representative cadre of experts. Given the litany of concerns about AI research, development, and deployment favoring countries in the Global North,[133] inadequate engagement with Global South values, perspectives, and experts would minimize the actionability of any research by the initiative. Recruiting a global set of AI experts, though, may be putting the sled before the dog.

Certain nations have yet to develop a cadre of AI experts who could potentially join an AI research initiative. For instance, efforts are underway to increase the number of students pursuing AI-related graduate degrees in Africa to bolster the continent's expertise in the area.[134] Similarly, AI stakeholders in Latin America and South America fear that a lack of AI expertise may partially explain why their representatives have had little sway in AI governance talks.[135] Even countries that have nationals with AI expertise may find that those experts move to countries with a higher concentration of experts—a brain drain dynamic that might make those expat experts less suited to represent the interests of their home country in an international institution.[136]

It follows that the international AI risk research initiative may have to develop more intentional processes to identify a global set of AI experts. The IPCC's approach, though increasingly nuanced in recent years, certainly does not guarantee a representative set of assessment authors: About 40 percent represented the Global South.[137] To improve on these numbers, the initiative may support efforts to foster more AI expertise around the world—perhaps by using some of its funds to invest in regional AI research centers.

---

[133] See, e.g., Robert Muggah & Ilona Szabó, "Artificial Intelligence Will Entrench Global Inequality," *Foreign Policy* (May 29, 2023).

[134] See, e.g., Eve Ruwoko, "5,000 PhD Scholars to Meet Africa's Growing AI Needs," *University World News* (Feb. 1, 2022).

[135] MIT Technology Review Insights, "The Global AI Agenda: Latin America," *MIT Technology Review* (June 8, 2020), https://perma.cc/4XK7-Z4KV.

[136] See, e.g., Karen Hao, "China's Path to AI Domination Has a Problem: Brain Drain," *MIT Technology Review* (Aug. 7, 2019), https://perma.cc/SGH2-AVRJ.

[137] Tandon, supra note 128.

The Power of Participation

In the same way economic, political, and cultural differences have thwarted the formation of a global climate institution with a broad regulatory mandate,[138] similar differences exist among AI stakeholders.[139] An inability to agree to a global regulator, though, need not stifle the creation of an international research organization. The IPCC came about because the international community had a shared desire for timely and authoritative assessments of the state of knowledge regarding climate change. A similar desire exists in the AI context.[140]

The IPCC tries to fill that knowledge gap in collaboration with national governments. Membership by the vast majority of nations, rather than a large budget or expansive regulatory mandate, contributes to the legitimacy of IPCC assessments as well as the influence of those assessments on climate regulation.[141] An alternative structure in which scientists alone governed the IPCC and authored its reports proved unviable. In particular, U.S. officials expressed "reluctan[ce] to cede epistemic authority on the issue [of climate research] to a purely international organization such as [the World Meteorological Organization], composed primarily of scientists." A similar reluctance would likely emerge if advocates for the responsible development and deployment of AI failed to provide national governments with some control over the research inputs and outputs. The extent to which this oversight helps or hinders the quality of IPCC science has long been a subject of discussion.

The role of IPCC assessments in sparking and directing international action on climate, however, suggests that the inclusion of national governments is a net positive. In other words, direct participation of the political actors intended to respond to the research in question may increase the odds of those actors taking such action. The same could be true in the context of an international AI risk initiative. Put differently, omitting national governments from an AI research effort may afford them more reasons to refuse whatever actions may be warranted by that research.

---

[138] UN General Assembly, "Resolution: Protection of the Global Climate for Present and Future Generations of Mankind" (1988) (endorsing the creation of the IPCC and tasking it with developing an overview of the state of knowledge of the science of climate change).

[139] See Huw Roberts, "A New International AI Body Is No Panacea," e-international relations, https://perma.cc/MT39-9XRJ; Rebecca Fannin, "In Tech Rivalry With the US, China Is Behind on a Key Asset: Its Own OpenAI," CNBC (March 31, 2024); see also Christina Montgomery, "Why a Single AI Regulator Just Won't Work," *The Hill* (May 25, 2023).

[140] See, e.g., Chris Vallance, "AI Risks Destabilising World, Deputy PM to Tell UN," BBC (Sept. 22, 2023).

[141] Paglia & Parker, supra note 110, at 299.

Final Recommendations for an International AI Risk Initiative

Enacting regulation that lacks grounding in the latest research is akin to sailing without a compass—you will move but likely not in the right direction. The R&R cycle prevents hasty and harmful regulation from being enacted or from staying on the books for too long. This paper makes the case for more attention being spent on the development of an international AI risk research initiative that can inform and incite regulation that is responsive to AI risks.

Though some advocates for the safe development of AI have hinted at the importance of AI risk research, many of these proposals have focused on state-sponsored research. Such proposals, though, often include limiting conditions with respect to the independence, comprehensiveness, and actionability of the underlying research. The National AI Research Resource under study by Congress, for instance, would omit the insights and values of a global set of AI scholars, would permit some of its finite resources to be used for commercial purposes, and would produce research unlikely to be accepted by certain countries. An international AI risk research initiative, in contrast, could mirror CERN or the IPCC by welcoming a global set of AI experts who, collectively, can conduct AI risk research that is more likely to be accepted and acted on by countries, AI labs, and other stakeholders.

At this stage in AI regulatory conversations, though, the important decision is not concluding whether a CERN for AI or IPCC for AI is more preferable; instead, the important decision facing AI stakeholders is whether to give AI risk research the attention it is due. This paper is strongly in favor of more resources being spent on the development of an AI research ecosystem than on the creation of the ideal AI regulator. If that research effort adheres to the three key lessons revealed by an examination of CERN and the IPCC, then it has good odds of kick-starting and sustaining the R&R cycle in the context of AI governance.

Lesson #1: Independent research is not by accident.

CERN has myriad structural, cultural, and legal aspects in place to facilitate independent, reliable research. The IPCC likewise has several safeguards in place to facilitate robust analysis of climate research. These deliberate and comprehensive approaches to directing research toward whatever the science demands deserve emulation by the AI initiative.

An international AI risk research initiative should copy and expand on the wise decisions made by CERN's founders and the creators of the IPCC. One particularly important step may be to ingrain a risk research focus into the initiative's DNA—for example, following the same approach as CERN by foreclosing any commercial research in its founding document.

Lesson #2: Resource-intensive research requires ongoing resource pooling.

CERN got up and running because several countries combined their resources and pledged to continue to do so on an annual basis. The IPCC operates solely because of the willingness of the world's brightest experts to consolidate their attention and knowledge.

The AI initiative should learn each part of this lesson: First, recognize that meaningful AI risk research will require substantial and continuous access to financial support; and, second, develop a network of supporters to ensure that the availability of such funds does not come into question. Of course, this latter lesson comes with some potential trade-offs—for one, inclusion of more financial supporters may complicate governance. Still, by virtue of anticipating that resource-pooling will be a core part of the AI initiative's success, supporters can build such pooling into the institution's structure and culture.

Lesson #3: Cutting-edge research benefits from global expertise.

Several of the founding CERN member states continued to operate national particle colliders despite the new, larger organization forming to replicate and improve on that work. In time, those national labs shut down. The demise of these smaller projects reflected the obvious superiority of doing particle research at scale—in terms of both infrastructure (i.e., a larger collider) and personnel (i.e., physicists). The consolidation of expertise in one location has facilitated the exchange of knowledge and resulted in tremendous cohesion within the physics community. This collaborative and productive research community publishes papers with hundreds of co-authors, conducts experiments in international teams, and attracts and retains the best and the brightest in the field.

Gathering AI experts at a single location a la CERN or directing them to resolve a specific research question a la the IPCC is a difficult task. Nevertheless, it can and must be pursued. First, though, the importance of an international, inclusive, and diverse research community needs to catch on among actors and individuals in the AI space. Alternatively phrased, exclusive subnational or national AI research efforts should be perceived as "back-up plans." Yet legislators from Sacramento to D.C. have discussed forming research resources accessible only to U.S. citizens or U.S.-affiliated organizations.

These lessons are three pebbles in a rockslide of good ideas available through study of CERN and the IPCC. I welcome and encourage more people to expand on this list. In the same way that the respective successes of CERN and the IPCC resulted from creative institutional thinkers implementing novel governance and operational mechanisms, the difficulties posed by launching an international AI risk research initiative will likely not be resolved by looking at tired and outdated concepts.

The Digital Social Contract paper series is supported by funding from the John S. and James L. Knight Foundation and Meta, which played no role in the selection of the specific topics or authors and which played no editorial role in the individual papers.