Washington Journal of Law, Technology & Arts

Volume 19 | Issue 3 Article 4

6-17-2024

THE NEED FOR AN INTERNATIONAL AI RESEARCH INITIATIVE: HOW TO CREATE AND SUSTAIN A VIRTUOUS RESEARCH-REGULATION CYCLE TO GOVERN AI

Kevin Frazier

WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS
VOLUME 19, ISSUE 3 — SUMMER 2024

THE NEED FOR AN INTERNATIONAL AI RESEARCH INITIATIVE: HOW TO CREATE AND SUSTAIN A VIRTUOUS RESEARCH-REGULATION CYCLE TO GOVERN AI

Kevin Frazier[1]

ABSTRACT

This paper explains the need for an international AI research initiative. The current focus of lawmakers at the subnational, national, and international level on regulation over research has created an imbalance, neglecting the critical role of continuous, informed research in developing laws that keep pace with rapid technological advancements in AI.

The proposed international AI research initiative would serve as a central hub for comprehensive AI risk analysis, modeled on successful precedents like CERN and the IPCC. CERN exemplifies a collaborative research environment with pooled resources from member states, leading to significant advancements in particle physics. Similarly, the IPCC has successfully consolidated and synthesized global climate research, informing policy decisions on an international scale. Drawing from these models, the initiative aims to provide accurate, timely assessments of AI risks, aiding policymakers worldwide and ensuring that AI development benefits all of humanity, not just technologically-advanced nations.

This paper also highlights the dichotomy in AI risk perspectives—near-term concerns like algorithmic bias versus existential threats like the empowerment of authoritarian regimes. This division often detracts from a unified approach to funding and researching all potential AI risks comprehensively. The necessity for an international body becomes evident as individual nations and private entities tend to focus on regional and domestic agendas, which are insufficient to address the global nature of AI risks.

Discussing various national and subnational efforts, the paper critiques their limited scope and emphasizes the inadequacies of these isolated initiatives in tackling global AI challenges. Instead, it calls for an international approach that can leverage global expertise and resources more effectively, similar to CERN's resource pooling and the IPCC's consensus-driven research aggregation.

In summary, the paper argues for a shift in focus from predominantly regulatory efforts to a balanced approach where informed, well-researched guidelines shape global AI policies. This shift is crucial to developing a regulatory framework that is responsive to the rapid advancements and broad implications of AI technologies. By fostering a robust international research initiative, stakeholders can ensure that AI development is guided by comprehensive risk assessments and ethical considerations, promoting a safer, more equitable technological future.

---

[1] Kevin Frazier is an Assistant Professor at St. Thomas University College of Law and a Director of the Center for Law & AI Risk. He is a proud graduate of the UC Berkeley School of Law (J.D.), Harvard Kennedy School (M.P.A.), and University of Oregon (B.S.). Professor Frazier clerked for Chief Justice Mike McGrath of the Montana Supreme Court following graduation. He has been conducting AI governance research with the Institute for Law and AI since early 2023. This article benefited from excellent research by Ana Barreto and commentary from Alan Rozenshtein and David Rubenstein. The Washington Journal of Law, Technology & Arts editorial team made this piece stronger as a result of their thoughtful edits and suggestions.

TABLE OF CONTENTS

INTRODUCTION

A virtuous research-regulation cycle is critical to mitigating risks from emerging technologies. In the AI context, though, stakeholders concerned about near- and long-term risks have focused more on regulation than research. That imbalance must come to an end. Absent the creation of an international AI research initiative that continuously evaluates AI risks, laws governing AI may become outdated or, worse, result in detrimental changes to the technology itself.[2]

AI risks are not easily explained, quantified, or mitigated. As a matter of fact, 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 or x-risks such as AI empowering and entrenching totalitarian regimes or enfeebling humans by rendering them over reliant on AI systems.[3] 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 with the capacity to mitigate the most likely and most severe risks identified through rigorous experimentation.

To the extent stakeholders, specifically public authorities such as EU member states, the United Kingdom, and the United States, as well as private actors including research organizations like the Ada Lovelace Institute and AI labs, have considered the need for AI risk research, they have generally prioritized domestic or regional responses over the development of an international institution. Such an institution would provide timely and accurate AI risk analysis for both the countries leading AI development as well as for those countries most susceptible to the harms associated with AI risks. Policymakers must reexamine that allocation of research and regulatory energy. Subnational or national research entities lack the funds as well as the legal and social mandate to meet a global need for AI research.

This paper explains why risk research is critical to any regulatory regime, proposes an international AI research initiative tasked with conducting AI risk research at scale. The two key examples used are CERN—"the undisputable hub of a global network of scientists without peer"[4]—and the Intergovernmental Panel on Climate Change (IPCC)— "the international body that reviews and assesses the latest science on climate change"—as models for how best to design and sustain such an initiative.

A CERN for AI would act as a centralized research hub that relies on resource-pooling by member states and other stakeholders to obtain sufficient compute and expertise to conduct its own AI risk. Such an institution would likely generate more timely and comprehensive assessments of AI risk. However, the low odds of developing such a hub warrant considering alternative research models. One such alternative is an IPCC for AI. This approach would involve recruiting AI experts to conduct periodic assessments of the latest AI risk research with the goal of reaching consensus as to the likelihood and severity of different AI risks. By

---

[2] *Ossification*, HTTP/3 EXPLAINED, https://http3-explained.haxx.se/en/why-quic/why-ossification (last visited Sept. 17, 2023).

[3] *See* 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), https://time.com/6303127/ai-future-danger-present-harms/.

[4] 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 211, 219-27 (Arjen Boin & Lauren A. Fahy eds., 2021).

consolidating, verifying, and summarizing pre-existing research, an IPCC for AI would require comparatively fewer funds and a smaller operation than an entity tasked with conducting its own audits and evaluations of AI models. Both research models would nevertheless mark an improvement on the status quo by providing the international community with a common, reliable, and accurate source of information on AI risks.

The goal of this paper isn't to advocate for the emulation of one existing organization over the other. Instead, it has two aims: first, to convince readers of the superiority of an international approach to AI risk research (and, thus, the need to prioritize the creation of an international AI risk research initiative); and, second, to nudge AI stakeholders to perceive CERN, the IPCC, and other organizations as sources of inspiration rather than as models to replicate. A successful risk mitigation regime must reflect the nature of the risks at issue. AI poses evolving, global, and irreversible risks that society has yet to confront, let alone mitigate. This paper does not identify the "right" approach to AI risk research nor dive into the technical aspects of that research. By pointing out distinguishing and exemplary traits of CERN and the IPCC, though, it may help stimulate and direct necessary conversations about an international AI research initiative.

PART I: THE VIRTUOUS RESEARCH-REGULATION CYCLE

The successful regulation of emerging technologies hinges on the development of a virtuous research-regulation cycle. The cycle operates by first conducting verifiable, reliable, and timely research on the risks posed by a new technology. Second, that research is distilled into actionable and understandable suggestions for regulation and presented to policymakers by neutral, authoritative actors. Third, those policymakers pass responsive regulation that incorporates the findings and, crucially, invests in a new wave of research to address unresolved questions.

The simplicity of the research-regulation cycle disguises the difficulty of creating and maintaining it. Where the cycle has emerged, though, society has benefited from substantial reductions in the risks posed by new technology. 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 IIHS and NHTSA then develop crash tests based on those accident reports to study and pinpoint weaknesses and flaws in automobile designs.[5] Those tests are conducted pursuant to a publicly available protocol and the results are made widely available in understandable formats.[6] Using these results, regulators update safety standards as well as crash reporting requirements.[7] For the purposes of this paper,

[5] *Ratings*, NHTSA, https://www.nhtsa.gov/ratings (last visited Sept. 17, 2023); *About our tests*, IIHS, https://www.iihs.org/ratings/about-our-tests (last visited Sept. 17, 2023).
[6] *About our tests*, IIHS, https://www.iihs.org/ratings/about-our-tests (last visited Sept. 17, 2023).
[7] *See* Second Amended Standing General Order 2021-01, NHTSA (Apr. 5, 2023), https://www.nhtsa.gov/sites/nhtsa.gov/files/2023-04/Second-Amended-SGO-2021-01_2023-04-05_2.pdf.

it is important to note that this instance of the cycle developed over decades in an emergent, rather than intentional, fashion.[8]

Climate change regulation provides another example of the research-regulation cycle in action. In this case, research is conducted by independent entities around the world before being verified, consolidated, and summarized by the IPCC.[9] Though the IPCC does not conduct its own research, it encourages scientists to conduct climate experiments by offering them an avenue to have their work evaluated and incorporated into authoritative assessments that are referenced by policymakers the world over.[10] Those policymakers then have the information necessary to craft new policies and to justify further investment in climate research.

Again, this cycle seems straightforward, yet the formation of the IPCC was not inevitable. Whereas direct, significant, and short-term financial interests of private and public actors in better understanding automobile safety facilitated the formation of the IIHS,[11] the IPCC aimed to fill a research gap with diffuse and uncertain public and private benefits. The development of IPCC's institutions, processes, and, perhaps most importantly, reputation was no walk in the park. The IPCC's formation took several years and its reports initially lacked the same influence and authority as contemporary versions.[12] Moreover, the actual work of the IPCC—creating assessments of the state of climate research—is "no easy task,"[13] especially when compared to ramming cars into walls. It takes IPCC researchers a tremendous amount of time, resources, and expertise to identify "consensus between its member governments and thousands of scientists and experts[.]"[14] The upshot is that research done primarily for public ends is unlikely to simply manifest, as such research requires extensive resources and planning.

The imminent risks posed by AI mean that policymakers cannot wait for the virtuous research-regulation cycle to arise naturally. An intentional effort to establish the processes and institutions essential to create the cycle in the AI context should coincide with efforts to develop regulation and inform those regulations. If one component of the cycle develops without sufficient consideration of the other, then neither will maximize its potential contribution. Imagine, for instance, that Congress passes a bill that inadequately funds AI research or omits a mechanism for AI labs to disclose certain critical information for such research. Amendment of that legislation may take months or years. Alternatively, attempts to correct such a mistake through rulemaking or other administrative procedures may be subject to significant legal risks.[15] Research must inform regulation and regulation must support research.

Notably, some policymakers appear to have caught on to the importance of coordinating the creation of AI regulatory and research institutions. California State Senator Scott Wiener, for

---

[8] *See* Kevin Frazier, *The Case for Prioritizing the Creation of an AI Benchmarking Consortium*, LAWFARE (Sept. 5, 2023), https://www.lawfaremedia.org/article/the-case-for-prioritizing-the-creation-of-an-ai-benchmarking-consortium.

[9] *About the IPCC*, THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, https://www.ipcc.ch/about/ (last visited Sept. 17, 2023).

[10] 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, 295, 301-02 (Arjen Boin eds., 2021).

[11] Kevin Frazier, *New Technology Requires New Regulatory Ambitions*, THE REGULATORY REVIEW (Oct. 9, 2023), https://www.theregreview.org/2023/10/09/frazier-new-technology-requires-new-regulatory-ambitions/.

[12] PAGLIA & PARKER, *supra* note 10, at 295-99.

[13] *Id.* at 295.

[14] *Id.*

[15] *See, e.g.*, TODD GARVEY, CONG. RSCH. SERV., A BRIEF OVERVIEW OF RULEMAKING AND JUDICIAL REVIEW 13, 17 (Mar. 27, 2017), https://sgp.fas.org/crs/misc/R41546.pdf.

instance, recently introduced a bill that called for a research cloud hosted by the state—"CalCompute"—that, as summarized by TIME, would "provide the computing infrastructure necessary for groups outside of big industry, like academia and startups, to do advanced AI work."[16] As discussed below, though, the development of anything other than an international research entity should be a secondary priority for AI stakeholders concerned about the manifestation of AI risks wherever they may occur.

PART II: WHY AI RESEARCH REQUIRES AN INTERNATIONAL ENTITY

The appropriate institutional setting for a specific research endeavor depends on the necessary inputs and the ideal outcomes of that research.[17] For example, high-energy particle collision experiments like those conducted at CERN require pooling of expertise and finances.[18] In contrast, certain forms of content moderation research can occur by pairing a few researchers with a single platform to conduct an experiment over a short time horizon.[19]

AI research is more like the former than the latter primarily because the quality, comprehensiveness, and authority of AI research improves with scale across many dimensions. Domestic research efforts may have to appease political actors by, for example, generating short-term, positive economic outcomes. Research into the full scope of risks posed by AI may or may not produce those results. Domestic research may also rely on a limited set of experts, reducing the likelihood of such research being as robust as possible. Finally, domestic researchers may lack the contextual information to map their findings onto other settings. In contrast, international research entities may develop *sui generis* norms and institutional structures that would be unlikely to form or perhaps even impossible to set up under the laws of a subnational unit or nation.

This part begins with a brief summary of two proposals for domestic AI research entities before detailing why the nature of AI risks demands an international AI research initiative.

A. DOMESTIC AI RISK RESEARCH PROPOSALS

Senator Wiener's proposal as well as a National AI Research Resource pilot program (NAIRR) under consideration by Congress are laudable and worthy of additional study. Though both are works in progress,[20] what information has been made available suffices to identify the high-level aspirations, institutional arrangements, and financial resources that would characterize

---

[16] Billy Perrigo, *Exclusive: California Bill Proposes Regulating AI at State Level*, TIME (Sept. 13, 2023, 7:00 AM), https://time.com/6313588/california-ai-regulation-bill/.

[17] *See* PAGLIA & PARKER, *supra* note 10, at 303 (discussing the importance of a research institution operating at a scale that matches the scope of the effects of the subject under study); Paul C. Stern, *Design principles for global commons: natural resources and emerging technologies*, 5 INT'L J. OF THE COMMONS 213, 219 (2011).

[18] *See, e.g.*, John Krige, *CERN from the mid-1960s to the late 1970s, in HISTORY OF CERN* 3, 16 (1st ed., 1986) https://medicalmarcom.com/medical-device-marketing/CERN-capitulo-1.pdf.

[19] *See, e.g.*, Jisu Kim, Curtis McDonald, Paul Meosky, Matt Katsaros, & Tom Tyler, *Promoting Online Civility Through Platform Architecture*, 1 J. OF ONLINE TR. & SAFETY (2022).

[20] *Id.*; News Release, Cal. Sen. Scott Wiener, *Senator Wiener Introduces Safety Framework in Artificial Intelligence Legislation* (Sept. 13, 2023) (on file with author), [hereinafter, *Wiener Press Release*].

them both. An initial overview of those high-level features helps illustrate their shortcomings and the need to prioritize an international AI research initiative.

As detailed below, the proposed international AI research initiative has several advantages over similar subnational or national research entities. The chief advantage is the initiative's intended focus on basic or pure research. This sort of research intends only to further a specific body of scientific knowledge, regardless of whether it will serve a direct commercial or practical end.[21] Given the uncertainty around how, when, and where AI's risks will manifest, such pure AI risk research is of immense social value,[22] though unlikely to be the subject of sufficient private funding.[23] Additional advantages include the autonomy of researchers to control their research agenda, the ability and authority to attract a global set of researchers, the financial security to procure and make available the necessary research tools, the governance stability to avoid the funding and decisional volatility associated with operating in a partisan environment, and the diversity of backgrounds and expertise among staff to mitigate concerns about the reliability of the research.

Before exploring those advantages, it's important to clarify the likely attributes of domestic AI research entities using the NAIRR as a proxy.

Stanford University Human-Centered AI (HAI) proposed a NAIRR to facilitate research by academia, government, industry, and civil society users.[24] These users would access "high-end computational resources, large-scale government datasets in a secure cloud environment, and necessary expertise[.]"[25] A congressionally-created task force tasked with studying the creation of a NAIRR envisions the research entity serving several purposes, such as encouraging commercial innovation and advancing the U.S.'s geopolitical interests.[26] The task force's final report, for example, mentions the importance of maintaining "American dominance" in AI R&D.[27] So long as that American-centric focus is at the core of a research endeavor, it is unlikely that stakeholders around the world will readily accept the reliability of such research.[28]

---

[21] Sirajul Islam & Sofiah Samsudin, *Basic Research and Its Importance to Enhance Fundamental Knowledge in Resources and Environment Friendly Technological Advancement: The Holy Qur'anic Prescription*, 10 INT'L J. SCI. & RES. PUB. 742, 743 (2020); Lord Rothschild, *Pure and Applied Research*, 120 J. ROYAL SOC. ARTS 205, 206 (1972) ("Basic research is done solely to increase knowledge.").

[22] Kelsey Piper, Thousands of AI experts are torn about what they've created, new study finds, VOX (Jan. 10, 2024), https://www.vox.com/future-perfect/2024/1/10/24032987/ai-impacts-survey-artificial-intelligence-chatgpt-openai-existential-risk-superintelligence.

[23] *See* Juan-Manuel Schvartzman & Jorge-Bernardo Schvartzman, *How do we ask for money? A view of funding for basic research*, 9 EMBO REP. 216, PASSIM (2008), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2267380/ (discussing efforts to find financial support for basic research).

[24] Stanford University Human-Centered A.I., *National AI Research Resource*, STAN. UNIV. HUMAN-CENTERED ARTIFICIAL INTELLIGENCE, https://hai.stanford.edu/policy/national-ai-research-resource (last accessed Sept. 17, 2023) [hereinafter, *Stanford NRC*].

[25] *Id.*

[26] NAT'L A.I. RSCH. RES. TASK FORCE, STRENGTHENING AND DEMOCRATIZING THE U.S. ARTIFICIAL INTELLIGENCE INNOVATION ECOSYSTEM, (Jan. 2023), https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf [hereinafter, *NAIRR Report*].

[27] *Id.* at 3.

[28] *Cf.* Richard Wike, Janell Fetterolf, & Mara Mordecai, *U.S. Image Plummets Internationally as Most Say Country Has Handled Coronavirus Badly*, PEW RESEARCH CENTER (Sept. 15, 2020), https://www.pewresearch.org/global/2020/09/15/us-image-plummets-internationally-as-most-say-country-has-handled-coronavirus-badly/ (analyzing survey results in which international respondents tended to lack trust in the United States, in part because of its political leadership and handling of the COVID-19 pandemic).

In short, a NAIRR would first and foremost be a U.S. research entity. It would operate pursuant to U.S. interests and exclusively serve users based in the U.S. or affiliated with U.S. organizations.[29] Moreover, governance decisions around the use of a NAIRR would incorporate "interests and perspectives from across Federal agencies."[30] Among those decisions would likely be the substance and enforcement of a user code of conduct.[31]

Finally, a NAIRR would rely on continued congressional funding to sustain its (costly) operations. To the task force's credit, it set forth a comprehensive budget to ensure users would receive access to state-of-the-art AI research infrastructure.[32] Still, even assuming the task force received its full budget request,[33] the resulting NAIRR would serve a relatively small user community of only about 19,000 users working on around 2,300 projects.[34]

B. WHY THE NATURE OF AI RISK RESEARCH REQUIRES AN INTERNATIONAL ENTITY

Neither CalCompute nor a NAIRR is a "bad" idea. This paper instead argues that those efforts are insufficient with respect to generating the research—and, by extension, regulation—necessary to mitigate AI risks wherever they manifest. The next section outlines the need for an international AI research initiative through two inquiries: first, an examination of the inputs, outputs, and outcomes associated with AI risk research; and second, an investigation of how best to mitigate the harms associated with "common-pool hazards,"[35] such as the deployment of AI models with the potential to wreak havoc.

1. The Inputs of AI Risk Research Necessitate Scale

Amassing the inputs necessary for cutting-edge AI research requires resource pooling. Scarcity plagues three of the most important inputs of AI research: data, compute, and expertise. On data, labs have used large amounts of data to increase the speed with which their models can learn.[36] It follows that researching AI models would likely benefit from having access to more

---

[29] NAIRR REPORT, *supra* note 26, at v.
[30] *Id.*
[31] Brendan Bordelon, *On AI, the government gets ready to throw its weight around*, POLITICO (May 16, 2023), https://www.politico.com/news/2023/05/16/the-government-plots-its-ai-approach-00097262.
[32] NAIRR REPORT, *supra* note 26, at 48.
[33] *See, e.g.*, Alan Rappeport, *U.S. National Debt tops $33 Trillion for First Time*, THE N.Y. TIMES (Sept. 18, 2023), https://www.nytimes.com/2023/09/18/us/politics/us-national-debt.html ("The debate over the debt has grown louder this year, punctuated by an extended standoff over raising the nation's borrowing cap.")
[34] NAIRR REPORT, *supra* note 26, at 48.
[35] *See* STERN, *supra* note 17, at 225 (explaining that common-pool hazards manifest when actors add "undesirable" bads to a common resource).
[36] *See* John Etchemendy & Fei-Fei Li, *National AI Research Resource: Ensuring the Continuation of American Innovation*, STAN. UNIV. HUMAN-CENTERED ARTIFICIAL INTELLIGENCE (Mar. 28, 2020), https://hai.stanford.edu/news/national-research-cloud-ensuring-continuation-american-innovation; *Fact Sheet: National Secure Data Service Act Advances Responsible Data Sharing in Government*, DATA COALITION (May 13, 2021), https://www.datacoalition.org/fact-sheet-national-secure-data-service-act-advances-responsible-data-sharing-in-government/.

data. Such a task is made easier by the participation of many actors, especially governments with significant amounts of data that is not yet accessible to private AI labs.[37]

On compute, the most advanced labs have levered better computational capacity and compute capability to create state-of-the-art AI models.[38] Quality AI research requires similar computational resources. 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] 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 Google annually spends $100 billion on its tech infrastructure.[40] Comparatively, the task force evaluating a NAIRR envisions an upfront investment of $2.6 billion to be supplemented annually with a little less than $1 billion.[41]

This limited budget may quickly leave users of a NAIRR in a lurch. GPUs—one basis to increase compute—are expected to experience yet another price spike in the near future.[42] And, if history repeats, then large AI labs may spend billions on whatever units are produced.[43] Resource pooling, then, is essential for research institutions to obtain sufficient 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 substantial or national entity.

And, if no research entity forms with the intent of making compute available to an international base of experts, then countries with insufficient resources to procure such compute will depend on compute-rich jurisdictions like California to do "pure" research on AI risks on their behalf or, even less likely, to grant their researchers access to the state-based resource. A return to Senator Wiener's bill and the proposed NAIRR shows why this arrangement falls short of the research required to help mitigate AI risks wherever it threatens harm. In particular, there's the timing concern; whereas an international research initiative could leverage its funds and broad range of support to quickly amass compute, the earliest the Wiener proposal could become law is the beginning of 2025.[44] A NAIRR would also take several years to get off the ground.[45] And, as mentioned above, the sustainability of such resources remains in question as the amount of funding required to keep pace with industry innovations continues to skyrocket.

---

[37] Daniel E. Ho et al., *Building a National AI Research Resource: A Blueprint for National Research Cloud*, 1 THE STANFORD INSTITUTE FOR HUMAN-CENTERED ARTIFICIAL INTELLIGENCE, 1, 9 (2021).
[38] *Id.*
[39] *Id.*
[40] Eze Vidra, *The Big Challenge for Generative AI Is GPU Capacity and Server Costs*, VC CAFE (Apr. 17, 2023), https://www.vccafe.com/2023/04/17/the-big-challenge-for-generative-ai-is-gpu-capacity-and-server-costs/.
[41] NAIRR REPORT, *supra* note 26, at 49.
[42] 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), https://finance.yahoo.com/news/gpu-prices-could-spike-again-052844387.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFh15xclKdbJ2yt3nduzo_c_MsP4HllXeXXCBCKiFIwTti-6BoJznBc09G-j0Lnr1DaVBK_yN8POghmtqPiwR5Zh7wXftSBzGIYDWas0hQekDF1MSvjMsrDoQL63kOsoTVudueyyGSSRInvISdLimsiwgdgyWVtd8zF2-__pHuXb.
[43] *Cf.* Kyle Wiggers, META bets big on AI with custom chips - and a supercomputer, TECHCRUNCH (May 18, 2023), https://techcrunch.com/2023/05/18/meta-bets-big-on-ai-with-custom-chips-and-a-supercomputer/ (flagging that Meta spent billions on NVIDIA GPUs but appears to be exploring an alternative means to boost compute).
[44] PERRIGO, *supra* note 16.
[45] NAIRR REPORT, *supra* note 26, at vi.

On expertise, no country has a monopoly on AI talent.[46] A review of AI-related publications across countries makes clear that expertise resides all across the world. 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.[47] 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.[48] Singapore's investments in its AI talent merits specific attention. Within the span of just three years, Singapore expects that USD $20 million worth of scholarships for the study of AI will triple its AI workforce.[49]An exclusionary or insular approach to AI risk research forecloses the possibility of bringing in 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. Just as many countries share an interest in expanding their respective AI talent pools, they also share a challenge: directing AI experts to jobs outside of industry. AI labs have successfully lured experts away from the public sector, specifically members of academic faculties.[50] 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 expertise in the field.[51] The aggregation of 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 "pure" research roles but also ensure that the technical resources made available at the international AI research initiative are being used to their maximum potential.[52]

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[.]"[53] 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. Senator Wiener admitted as much upon introducing his CalCompute bill, noting the limited

---

[46] *See* Neil Savage, *The race to the top among the world's leaders in artificial intelligence*, NATURE (Dec. 9, 2020), https://www.nature.com/articles/d41586-020-03409-8 (discussing rates of AI-related publication by country).
[47] China AI Development Report 2018, CHINA INSTITUTE FOR SCIENCE AND TECH-NOLOGY POLICY AT TSINGHUA UNIVERSITY, 12, (July 2018).
[48] Udit Sabharwal et al., *Artificial Intelligence Tech Hubs: Asia Pacific Talent Spotlight*, CBRE (May 23, 2023), https://www.cbre.com/insights/articles/artificial-intelligence-tech-hubs-asia-pacific-talent-spotlight.
[49] *Press Release: Artificial Intelligence (AI) initiatives launched to uplift Singapore's economic potential*, MINISTRY OF COMMUNICATIONS AND INFORMATION (Mar. 1, 2024), https://www.mci.gov.sg/media-centre/press-releases/ai-initiatives-launched-to-uplift-sg-economic-potential/.
[50] Brian Eastwood, *Study: Industry now dominates AI research*, MIT (May 18, 2023), https://mitsloan.mit.edu/ideas-made-to-matter/study-industry-now-dominates-ai-research.
[51] Remco Zwetsloot & Jack Corrigan, *AI Faculty Shortages*, CSET 5 (July 2022) ("[W]e identified a variety of indirect evidence that suggests universities are struggling to meet students' growing demand for AI education.").
[52] Cameron F. Kerry et al., *Strengthening international cooperation on AI*, BROOKINGS (Oct. 25, 2021), https://www.brookings.edu/articles/strengthening-international-cooperation-on-ai/#:~:text=AI%20research%20and%20development%20is,comparative%20advantages%20for%20mutual%20benefit.
[53] *New McKinsey survey reveals the AI tech-talent landscape*, MCKINSEY & CO. (Jan. 20, 2023), https://www.mckinsey.com/about-us/new-at-mckinsey-blog/ai-reinvents-tech-talent-opportunities.

capacity of the California state government to audit AI systems and otherwise implement the provisions of his bill.[54] 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.

1. The Outputs of AI Risk Research Support an International Approach

AI research should produce objective, comprehensive, and timely analysis of the capacity of existing and soon-to-be deployed AI models as well as the risks posed by those models. This information can and should steer policy development at the international, national, and subnational levels.[55] Yet, if a subnational or national government leads such research, the odds of the outputs being objective and trustworthy diminish, the research becomes less comprehensive, and timeliness of the research may suffer. Insufficient access to the aforementioned inputs only explains part of these limitations. Another reason why an international AI research initiative is better suited to produce such outcomes is its exclusive focus on "pure" research.

Proposed research institutions with a direct relationship to a subnational or national government often involve some "developmental" goal. Senator Wiener's press release on his proposed legislation demonstrated this dynamic. He explicitly stated that the CalCompute system would be used to "foster innovation for small businesses[.]"[56] More generally, the Senator indicated a desire to assist with the development of AI, noting that 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.[57] Relatedly, the proposed NAIRR, though intended for use by academic and non-profit researchers, also has a developmental purpose. 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."[58]

Commercial concerns and a focus on nationalistic competition 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 the research entity's resources, for instance, may not want to disclose certain inputs and outputs, and the same is true 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. For example, the Ada Lovelace Institute pointed out that OpenAI withheld certain information about GPT-4 upon its deployment.[59] An expectation of total transparency—at least with respect to the public—goes a step too far, as recognized by the Institute.[60] Perhaps more important than what OpenAI shared is the fact that it exercised complete discretion over its decision.[61] If the research entity relies on corporate investment or participation or has too much of a focus on commercialization, then such omission of

[54] PERRIGO, *supra* note 16 (quoting Senator Wiener).
[55] *See, e.g.*, Elliott Jones, *Keeping an eye on AI*, ADA LOVELACE INSTITUTE at 4 (July 2023) ("There are specific properties of AI systems that the Government should consider measuring and monitoring such as their deployment and governance, and their downstream impacts on individuals and society.").
[56] WIENER PRESS RELEASE, *supra* note 20.
[57] *Id.*
[58] STANFORD NRC, *supra* note 24.
[59] Jones, *supra* note 55, at 6, 9.
[60] *Id.*
[61] *Id.*

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 and ensuring 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.[62] Even citizens may refrain from joining such an entity if they thought their work could lead to controversial domestic or national security ambitions.[63]

With the proper governance structure and funding mechanisms, an international AI research initiative could exclusively conduct "pure" research with the participation of experts and AI labs located in a broad range of countries. Moreover, if the initiative manages to develop a reputation 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.[64] Finally, such an initiative—unconcerned with trade secrets and state secrets—could more freely disclose its research methodologies and findings. This transparency could aid the identification of any flaws or areas for improvement in the initiative's research.

2. The Intended Outcomes of AI Risk Research Warrant an International Approach

AI research must not be done only for the sake of knowledge. Given the aforementioned risks of AI, research should be designed and summarized to inform policymakers around the world and influence their adoption of certain regulations.

Just as AI research and development is taking place across the globe, the effects of such R&D will affect the lives of billions. Subnational and national governments cannot avoid the ongoing and complex regulatory challenge of steering AI development towards the interests of their citizens. Absent the creation of an international AI research initiative, though, some governments may not have access to reliable and actionable guidance. Importantly, these may be the countries most in need of guidance on the most effective regulations to mitigate AI risks.[65]

Even the most well-intentioned research entity will fall short of its ability to inform decision making around AI regulation if it is not inclusive of the global community. The creation, deployment, and regulation of emerging technology has traditionally been dominated by nations in the Global North. Yet, such actions have effects on the people and governments in the Global South.[66] If the national membership and staff of a research entity appears to perpetuate this

---

[62] Exec. Order 11935, 41 FED. REG. 37301 (Sept. 3, 1978) (placing limits on federal government hiring of non-citizens).

[63] Emily Badger, Quoctrung Bui &Alicia Parlapiano, *The Government Agencies That Became Smaller, and Unhappier, Under Trump*, THE N.Y. TIMES, https://www.nytimes.com/2021/02/01/upshot/trump-effect-government-agencies.html#:~:text=By%202020%2C%20there%20were%20300,at%20the%20Department%20of%20Labor (updated Oct. 13, 2021).

[64] ARJEN BOIN, LAUREN A. FAHY & PAUL 'T HART, GUARDIANS OF PUBLIC VALUE: HOW PUBLIC ORGANIZATIONS BECOME AND REMAIN INSTITUTIONS, IN GUARDIANS OF PUBLIC VALUE: HOW ORGANIZATIONS BECOME AND REMAIN INSTITUTIONS, 1, 6 (Arjen Boin et al. eds., 2021).

[65] Robert Muggah Gabriella Seiler & Gordon LaForge, *AI and the Global South*, PROJECT SYNDICATE (Mar. 2, 2023), https://www.project-syndicate.org/commentary/ai-governance-first-principles-must-include-global-south-by-robert-muggah-et-al-2023-03.

[66] Mark Scott, *The Global South's missing voice in AI*, POLITICO (Aug. 31, 2023, 1:30 PM), https://www.politico.eu/newsletter/digital-bridge/the-global-souths-missing-voice-in-ai/.

historical inequality, then the odds of its research serving as the common basis for consistent regulatory actions will diminish.[67]

Hesitancy among Global South nations to accept any supposedly international endeavor—even one with an explicit mission to mitigate risks that transcend borders—likely stands at an all-time high following COVID-19. Though years have passed since the height of the pandemic, leaders such as President Cyril Ramaphosa of South Africa have not forgotten that, in his words, "The Northern Hemisphere countries . . . were hogging" vaccines and "didn't want to release them at the time when we needed them most."[68] If President Ramaphosa speaks for even a fraction of similarly-situated leaders, then an AI research entity will have to address the perception that leaders of Global North countries place less value on the lives of people located below the equator.[69] Global South experts and leaders may withhold their support of any international AI research initiative without clear indications that all countries will have a meaningful stake in its governance, operations, and outputs.

The upshot is that a research entity that intentionally and successfully attracts and fosters international participation should increase the odds of coordinated action at the international level. This is especially true if international consensus is reached on the interpretation and importance of the research entity's findings. Consensus around the likelihood and severity of different risks could set off another, different virtuous cycle: identification of a priority risk area could direct more research into that area as well as increase willingness among stakeholders in AI research, development, and deployment to consider international action to combat that risk. Any resulting success in the mitigation of that risk could serve as validation of the initiative's mission and perpetuate its contribution to this specific research-regulation cycle.[70]

3. Other Benefits of an International Approach to AI Risk Research

An international AI research initiative designed pursuant to the best practices set forth by similar organizations, such as CERN and the IPCC, can exercise institutional flexibility more so than subnational or national bodies. The latter entities may face political, financial, or legal constraints on their ability to adapt to new conditions and tasks. In other words, subnational and national research entities likely will embody the telltale signs of twentieth-century organizations: being "grounded in hierarchy, specialization, and compartmentalization of knowledge, funding, task performance and responsibilities."[71] Such constraints are inconsistent with the requirements of an AI research entity tasked with evaluating an unpredictable emerging technology. Such an entity needs the characteristics of successful institutions in the twenty-first century:

---

[67] *See, e.g.*, PAGLIA & PARKER, *supra* note 10, at 297 (summarizing criticism levied against the IPCC for "[o]ver-representation of knowledge produced in industrialized countries . . . .").
[68] Mark Suzman, *The Roots of the Global South's New Resentment*, FOREIGN AFFAIRS (Sept. 8, 2023), https://www.foreignaffairs.com/africa/roots-global-souths-new-resentment.
[69] *Id.*
[70] Boin et al., *supra* note 64, at 6.
[71] *Id.* at 30; *see, e.g.*, Lindsey Barrett et al., *Outdated Ethics Rules May be Stymieing the Federal Trade Commission's Efforts to Keep Up with Big Tech*, BROOKINGS (July 29, 2021), https://www.brookings.edu/articles/outdated-ethics-rules-may-be-stymieing-the-federal-trade-commissions-efforts-to-keep-up-with-big-tech/ (providing an example of how federal administrative agencies often struggle to adjust to new conditions due, in part, to their rigid structures and outdated rules); David McCabe, *How California is Building the Nation's First Privacy Police*, THE N.Y. TIMES (Mar. 15, 2022), https://www.nytimes.com/2022/03/15/technology/california-privacy-agency-ccpa-gdpr.html (detailing barriers to launching a new state agency in California).

"collaborati[on] across boundaries, pooling of resources, flexible arrangements, shared power and responsibilities."[72]

Analysis of the issues affecting federal research entities confirms that an institution freed from compliance with traditional governance structures would improve the institutional capacity of an AI research entity. For one, an international research initiative could develop a funding mechanism shielded from sways in political opinion. Comparatively, congressional spending panels led by Republicans recently proposed slashing the budget of President Biden's prized Advanced Research Projects Agency for Health from $.15 billion to $500 million.[73] Even when politics are not at the core of a budget decision, Congress has long indicated skepticism toward funding "pure" research.[74] In the early 1990s, the U.S. House killed a proposal to build a superconducting super collider to compete with CERN's Large Hadron Collider in part because representatives doubted the societal value of such research.[75] More recently, Congress has often placed research programs at the front of the list for cuts when looking to reduce federal spending.[76]

An international AI research initiative could, like CERN, adjust its governance structures and norms to "allow for its projects to originate from and be driven by the needs and ambitions of science[.]"[77] Comparatively, the U.S. Department of Energy (DOE)-operated national laboratories have found themselves "locked in an unproductive bureaucratic logjam."[78] A 1998 report by the Government Accountability Office confirmed as much, describing the nation's labs as "unfocused, . . . micromanaged by DOE, and [unable to] function as an integrated national research and development system."[79]

Though decades have passed, such problems have persisted. A 2017 report by the Harvard Kennedy School (HKS) noted "increasingly prescriptive management" of the labs.[80] The HKS team attributed that constraint on research to "mistrust between Congress, DOE, the Labs, and the broader industrial and scientific community[.]"[81] Cumulatively, these trends have driven up costs and prevented the effective operation of the labs.[82] Similar issues span the entire lab

---

[72] Boin et al., *supra* note 64, at 30.
[73] Jeffrey Mervis & Jocelyn Kaiser, *Congressional Spending Sanels Cruel to NIH, Kinder to NSF*, SCIENCE (July 14, 2023), https://www.science.org/content/article/congressional-spending-panels-cruel-nih-kinder-nsf.
[74] Richard D. McCullough, *The Lack of Funding Is a Tragedy for Bold Scientific Breakthroughs*, THE N.Y. TIMES (Sept. 16, 2016), https://www.nytimes.com/roomfordebate/2016/09/20/the-cost-of-corporate-funded-research/the-lack-of-funding-is-a-tragedy-for-bold-scientific-breakthroughs.
[75] Barry James, *Europe is Ready to Pick Up the Pieces in Particle Research*, THE N.Y. TIMES (June 20, 1992), https://www.nytimes.com/1992/06/20/IHT-europe-is-ready-to-pick-up-the-pieces-in-particle-research.html?searchResultPosition=98.
[76] *See, e.g.*, Madeleine Ngo, *CHIPS Act Funding for Science and Research Falls Short*, THE N.Y. TIMES (May 30, 2023), https://www.nytimes.com/2023/05/30/us/politics/chips-act-science-funding.html.
[77] ENGELEN & 'T HART, *supra* note 4, at 230.
[78] *Hearing Before the Subcomm. on Basic Research and Energy and Environment*, 105 CONG. (1998) (statement of Dr. Charles V. Shank, Director, Lawrence Berkeley National Laboratory), https://www2.lbl.gov/Science-Articles/Archive/CVS-Sept-23-98.html.
[79] U.S. Gen. Accounting Office, GAO-98-197, Department of Energy: *Uncertain Progress in Implementing National Laboratory Reforms* (1998), https://www.gao.gov/assets/rced-98-197.pdf.
[80] Amitai Y. Bin-Nun, Gabriel Chan, Laura Diaz Anadon, Venkatesh Narayanamurti & Sarah Jane Maxted, *The Department of Energy National Laboratories*, HARVARD KENNEDY SCHOOL, 35, 36 (Nov. 2017), https://dash.harvard.edu/bitstream/handle/1/37373236/enrp-stpp-lab-report-final-1.pdf?sequence=1&isAllowed=y.
[81] *Id.* at 36.
[82] *Id.*

system—in particular, labs have pursued duplicative research inquiries and seen their missions creep into overlapping areas.[83] Comparatively, a review of the procedural and structural changes adopted by CERN and IPCC indicate that, with the proper systems and governance structures in place, international research entities exhibit a significant degree of adaptability.

4. Validation of Prioritizing an International Approach to AI Risk Research

Unsurprisingly, the rationale for an international AI research initiative set forth above aligns with the arguments made by proponents of an international AI regulatory entity. OpenAI's leaders, for instance, called for the creation of an international watchdog akin to the International Atomic Energy Agency to regulate AI's development.[84] OpenAI's arguments for such a watchdog also bolster the case for an international approach to AI research. For instance, the ChatGPT creators emphasized the global nature of AI risks as well as the potentially existential threats posed by those risks as warranting a collaborative global effort.[85] They likewise noted that international collaboration can benefit nations that might not otherwise have the governance resources required to take on the tasks at issue.[86]

Other AI experts have echoed those arguments.[87] Paul Samson, for one, amplified his call for an international regulatory authority by noting that current regulatory "initiatives are scattered across nations and do not provide a foundation for a unifying regulatory framework upon which to build international AI regulation."[88] He also emphasized the importance of pursuing a regulatory effort that encapsulates a diverse set of values and cultural norms. This, Samson argued, has not been the case with respect to privacy regulation. In that context, Europe's proactive approach to regulation created a "Brussels Effect" that resulted in non-EU countries adopting aspects of the EU's approach in their own regulation. An undesirable and perhaps unintended byproduct of that embrace of EU policy is the spread of the EU's individualistic conceptualization of rights—a perspective that is not universally shared.[89] AI research and regulation can and should avoid being shaped by the values of a subset of the global community—an outcome that's only possible with the creation of international bodies.

Another argument for an international AI research and regulatory body reflects the fact that AI labs have so far demonstrated an unwillingness to impose meaningful self-regulations, according to Gary Marcus and Anka Reuel. OpenAI revealed this trend when, after acknowledging serious concerns with GPT-4, did not suggest nor take any responsive actions.[90] And, of those regulations that have been passed at the national level, "there is little global coordination."[91] This approach to regulation—whether best characterized as hands-off or

---

[83] *Id.* at 29, 51-52.
[84] Gregory Schmidt, *A.I. Needs an International Watchdog, ChatGPT Creators Say*, THE N.Y. TIMES, (May 24, 2023), https://www.nytimes.com/2023/05/24/business/artificial-intelligence-regulation-openai.html.
[85] Sam Altman, Greg Brockman & Ilya Sutskever, *Governance of superintelligence*, OPENAI (May 22, 2023), https://openai.com/blog/governance-of-superintelligence.
[86] *Id.*
[87] *See, e.g.*, Gary Marcus & Anka Reuel, *The world needs an international agency for artificial intelligence, say two AI experts*, ECONOMIST (Apr. 18, 2023), https://www.economist.com/by-invitation/2023/04/18/the-world-needs-an-international-agency-for-artificial-intelligence-say-two-ai-experts.
[88] Paul Samson, *On Advancing Global AI Governance*, CENTRE FOR INTERNATIONAL GOVERNANCE INNOVATION (May 1, 2023), https://www.cigionline.org/articles/on-advancing-global-ai-governance/.
[89] *Id.*
[90] MARCUS & REUEL, *supra* note 87.
[91] *Id.*

decentralized—may enhance risks posed by AI models. Marcus and Reuel argue, "An uneven, loophole-ridden patchwork [of AI regulations] is to no one's benefit and safety."[92] Likewise, research that turns on the willingness of labs to evaluate their own models may leave an incomplete and inaccurate picture of AI risks.

Observers have not, however, shied away from questioning the feasibility of such an international regulator—a pessimistic, albeit important, sentiment that should inform calls to create an international research initiative. In short, international regulatory bodies have traditionally formed in response to global catastrophe rather than in anticipation of one.[93] What's more, such bodies rarely form immediately after the disaster, further delaying an international response to the issues that gave rise to the event in question. The International Atomic Energy Administration, for instance, was established more than a decade after the U.S. used atomic weapons on the Japanese public.[94] Whether an international AI research initiative can form prior to the occurrence of a catastrophic AI event is unclear. Even if the will to form such an organization does not presently exist, the more an international AI research initiative is discussed, debated, and refined, the sooner it can come into being once international sentiment supports the idea.

In summary, a diverse and large set of actors acknowledge that an international regulatory entity has a clear role in the larger research-regulation cycle. Many of the justifications for that entity also make the case for an international research entity. The proactive creation of either, though, will require an unparalleled effort by a broad set of motivated and generous actors.

5. Theoretical Basis for an International AI Research Initiative

A more abstract analysis of the proper AI research and regulatory framework reinforces the prior practical analysis. Risks from emerging technologies are "common-pool hazards."[95] These types of problems are related to the more frequently discussed "tragedy of the commons." Degradation characterizes both problems but manifests differently in each. In the latter, degradation arises from the extraction of goods from a commons, while in the former, degradation occurs from the addition of undesirable "bads."[96] In other words, the developers and disseminators of potentially harmful emerging technologies do not reduce the supply of a resource but rather add to common-pool risks.

Both acts—subtracting from a commons and spoiling a commons—present similar regulatory challenges. These challenges are due to "complexity of choices, multidimensionality of risk, scientific uncertainty, value conflict and uncertainty, long time horizons, difficulty of excluding actors from taking action, high stakes potential for mistrust, and time pressure."[97] any of the strategies that scholars have identified as means to resolve "tragedy" problems—like climate change—may assist with mitigating "common-pool hazards" like the unchecked dissemination of emerging technologies.[98]

---

[92] *Id.*
[93] *See* Ryan Heath, *Don't hold your breath for global AI rules*, AXIOS (May 30, 2023), https://www.axios.com/2023/05/30/global-ai-regulation-catastrophe.
[94] *Id.*
[95] STERN, *supra* note 17, at 225.
[96] *Id.*
[97] *Id.* at 226.
[98] *Id.*

Broad remedial principles that apply in both contexts include: supporting scientific inquiries related to the "pool" at issue, creating independent monitoring authorities, providing users with meaningful opportunities to participate in the management of that pool, disseminating findings from scientific inquiries in a transparent, understandable, and actionable format, and involving a range of actors from local to global institutions.[99] It is worth noting that the parallels between mitigating climate change risks and AI risks may become stronger over time as the development, deployment, and use of AI increasingly exacerbates the underlying causes of climate change.[100] Climate researchers have particular concerns about the resource-intensive nature of AI development as well as the likelihood that its misuse will spread misleading information about climate science.[101]

The mitigation of common-pool hazards arising from emerging technology requires additional, specific governance responses based on distinctive aspects of such hazards. In particular, Paul Stern argues that technological innovation, more so than climate change, poses permanent harms. Whereas natural systems typically do not go past tipping points beyond which "they shift irreversibly into a new state," technological innovation generally results in irreversible effects.[102] This characteristic necessitates proactive risk identification and regulation. Of course, such anticipatory initiatives involve many complicating factors such as "scientific complexity, uncertainty, and ignorance[.]"[103] The alternative—waiting for the risks to become more apparent as the technology becomes more widely deployed and integrated—is unjustifiable given the irreversibility of the risks.

Moreover, whereas "traditional" commons issues are global in the sense that local decisions can alter integrated systems that span the globe, emerging technologies may cause direct harm anywhere. The transnational effects of emerging technologies are the product of two realities: technological expertise being easily shared, transferred, and emulated, and global markets facilitating the rapid and widespread distribution of technological innovation.[104] Consequently, because "anyone may be affected by technological risks, . . . everyone has a stake in [its] governance[.]"[105]

Though theory justifies the need for an international AI research initiative, it also explains why such an effort faces stiff odds of success. The public's limited understanding of AI may diminish their demand for such an initiative. No simple educational programs can address this reality. Not only do emerging technologies include complex topics, they also involve a distinct kind of existential risk that humans often struggle to accurately perceive and quantify.[106] For example the results from a survey of thousands of Germans about different risks posed by AI led researchers to "conclude that AI is still a 'black box' for many."[107] They theorized that the

---

[99] *Id.* at 224-25.
[100] Jude Coleman, *AI's Climate Impact Goes beyond Its Emissions*, SCIENTIFIC AMERICAN (Dec. 7, 2023), https://www.scientificamerican.com/article/ais-climate-impact-goes-beyond-its-emissions/.
[101] Oliver Milman, *AI likely to increase energy use and accelerate climate misinformation – report*, GUARDIAN (Mar. 7, 2024), https://www.theguardian.com/technology/2024/mar/07/ai-climate-change-energy-disinformation-report.
[102] STERN, *supra* note 17, at 226.
[103] *Id.*
[104] *Id.*
[105] *Id.* at 226-27.
[106] Paul Slovic, *Perception of Risk*, 4799 SCIENCE 280 (Apr. 17, 1987).
[107] Philipp Brauner, Alexander Hick, Ralf Philipsen & Martina Ziefle, *What does the public think about artificial intelligence?—A criticality map to understand bias in the public perception of AI*, 5 FRONTIERS IN COMPUT. SCI. (Mar. 16, 2023), https://www.frontiersin.org/articles/10.3389/fcomp.2023.1113903/full.

uncertainty around AI risk "can lead to biased and irrational control beliefs in the public perception of AI."[108] What's more, absent proactive research by a neutral, trusted institution, then ever more tangible AI risks, such as job elimination,[109] may not be studied and grokked by the public.

In the participatory gap left by a confused or ignorant public, technical experts with the requisite knowledge tend to shape risk regulation on behalf of general society. This substitution is imperfect. Regulatory decisions made by such experts "are not necessarily widely accepted as legitimate[.]"[110] Unsurprisingly, this class of experts—an unrepresentative subset of the population—may fail to account for backgrounds and perspectives of many communities around the world, especially those with comparatively less social and economic capital.[111] Stern maintains that this possibility necessitates a specific kind of governing institution with respect to emerging technologies. In particular, he argues such governance requires "institutions that ensure that the interests and values of all those facing common-pool hazards are involved in developing governance rules and in making decisions, even if most of those individuals cannot be meaningfully involved."[112]

This theoretical background underscores the need for an international AI research initiative. Such an initiative can address many of the governance difficulties associated with this common-pool hazard: it can enhance the public's understanding and appreciation of AI risks, facilitate the exchange of diverse viewpoints among a global set of experts, and assist with anticipatory governance by virtue of having the expertise and resources required to conduct research on the latest and most advanced AI models.

C. SUMMARY OF CASE FOR AN INTERNATIONAL AI RESEARCH INITIATIVE

This section argues that, if properly designed and maintained, an international AI research initiative could play a significant role in the virtuous research-regulation cycle described above. An international AI research initiative—as established by the inputs and intended outputs and outcomes of that research—would benefit from scale. For instance, such an entity could establish a set of diverse and well-endowed sources of funding so as to procure the latest inputs to conduct cutting-edge AI research. Likewise, such an entity could establish a global network of individual experts and smaller research institutions to evaluate new models in a timely and comprehensive fashion. Finally, an international entity could assist in achieving consensus about AI risks by virtue of being inclusive of a range of countries, stakeholders, and values.

An international research entity could also avoid capture by a subnational or national host government. This capture can manifest in several deleterious ways, including control over the research agenda by administrators or private entities contracted to run the labs,[113] micromanagement of research tasks, and refusal to alter the lab's structure or experiments based on changes in the underlying science or subject matter. On the whole, again from the U.S.

---

[108] *Id*.
[109] *Findings: The Potential Future Risks of AI*, NATIONAL ARTIFICIAL INTELLIGENCE ADVISORY COMMITTEE at 2 (Oct. 2023), https://ai.gov/wp-content/uploads/2023/11/Findings_The-Potential-Future-Risks-of-AI.pdf.
[110] STERN, *supra* note 17, at 227.
[111] *Id*.
[112] *Id*.
[113] BIN-NUN et al., *supra* note 80, at 41.

perspective, national labs have yet to attain the autonomy that a diverse range of think tanks have advocated for over several decades.[114]

The case for allocating a share—if not a significant share—of the U.S.'s AI-related expertise and resources to an international AI research initiative is further supported by a culture and set of institutional factors that "create challenges in funding capital-intensive initiatives and can actually encourage [the national laboratories] to become excessively conservative to preserve their portion of overall funding."[115] These concerns should give pause to advocates of prioritizing the creation of a state-based or federal AI research entity. Reliance on a patchwork of domestic research initiatives to mitigate AI's transnational risks seems like a bad bet. Research into the weaponization of AI and use of AI to deploy bioweapons, for instance, would benefit from an international perspective.[116]

At a time of scarce public AI research resources, both personnel and capital should go to the entity best able to contribute to the virtuous research-regulation cycle. If an international AI research initiative develops with the governance structure of CERN and consensus-building capacities of IPCC, then it will surely garner support from the U.S. and all other countries interested in a collaborative effort to mitigate AI risks. If and when a subnational unit or a nation has the institutional capacity to add to that cycle, then a greater portion of a nation's resources should flow there. In the interim, nations should collaborate to design and launch an international AI research initiative capable of realizing the comparative advantages of such an entity.

Finally, to the extent that subnational and national research entities are already in development or up and running, they should consider themselves part of an international network of such labs that collaborate and coordinate through the international AI research initiative. Without a central repository and consolidator of such research, nations lacking such research entities will have limited to no means of learning about the latest strategies to mitigate AI risks. Consequently, subnational and national labs already conducting research should take steps now to ensure they can share information and findings with a global entity, including having users agree to certain information sharing arrangements.

PART III: LESSONS FROM CERN AND THE IPCC ON FORMING AN INTERNATIONAL AI RESEARCH INITIATIVE

Though the need for and potential of an international AI research initiative are clear, what remains unclear is how best to organize such an entity. Some useful suggestions, however, arise from analysis of two science-based institutions with international stakeholders and broad mandates to conduct "pure" research: CERN and the IPCC.

CERN operates several cutting-edge particle accelerator facilities with the goal of "uncover[ing] what the universe is made of and how it works."[117] The organization emerged from a unique historical moment that enabled robust national cooperation and encouraged

[114] *Id.* at 31-32.
[115] *Id.* at 52.
[116] *See* Charles Ovink, *AI risks for international peace and security*, OBSERVER RESEARCH FOUNDATION (Jan. 31, 2024), https://www.orfonline.org/expert-speak/ai-risks-for-international-peace-and-security (calling for a multi-stakeholder, international approach to AI risk research).
[117] CERN, *Our Mission,* https://www.home.cern/about/who-we-are/our-mission (last visited Sept. 17, 2023).

substantial investment in "big" scientific endeavors.[118] Decades later, CERN serves as a model for how to break through the gridlock that, with increasing frequency, undermines international collaboration.[119] For the purposes of this article, 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.

The IPCC authors "authoritative reports [that] inform international policy and negotiations on climate change."[120] 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."[121] Particularly relevant here is the IPCC's creation of a global network of experts and implementation of trusted processes that allow member states to achieve consensus on complex and controversial scientific matters.

As public organizations, CERN and the IPCC have distinguished themselves as "institutions." Whereas an organization is "nothing more than an established way of cooperation between two or more individuals," according to Arjen Boin, Lauren Fahy, and Paul t' Hart, an institution sets its target on "aims that are widely considered to fulfill a societ[al] need," demonstrates "reliable performance over time," and performs in an "exemplary" manner from the perception of its societal constituencies.[122] Other such "institutions" include the Corrupt Practices Investigations Bureau in Singapore, the BBC, the European Court of Justice, the World Anti-Doping Agency, and Medecins Sans Frontieres. This article focuses on CERN and the IPCC because of the scientific nature of their work, their supranational constituencies and staff, and because AI governance stakeholders have already indicated a willingness to consider these institutions as models for an international AI entity.[123]

A. IDENTIFYING THE DISTINGUISHING TRAITS OF CERN AND THE IPCC

Calls for a "CERN for AI" or "IPCC for AI" are commonly as short on details as they are in length. Op-ed-length pieces on the topic often fail to explore the extent to which CERN and IPCC cannot nor should not be exactly replicated by an international AI research initiative. By way of example, an attempt to copy and paste CERN into the AI context would likely flounder for several reasons, chief among them an unwillingness among the international community to

---

[118] ENGELEN & 'T HART, *supra* note 3, at 214-18; Article II, § 1, CERN CONV. (July 1, 1953). https://council.web.cern.ch/en/content/convention-establishment-european-organization-nuclear-research#2.
[119] Mark Robinson, *The CERN Community; A Mechanism for Effective Global Collaboration?*, 10 GLOBAL POLICY 41 *PASSIM* (2019).
[120] PAGLIA & PARKER, *supra* note 10, at 295.
[121] *Id*. at 296.
[122] BOIN et al., *supra* note 64, at 3.
[123] *See, e.g.*, Georg Scholl, *We need a CERN for AI in Europe*, ALEXANDER VON HUMBOLDT STIFTUNG (Aug. 1, 2022), https://www.humboldt-foundation.de/en/explore/magazine-humboldt-kosmos/by-courtesy-of-how-artifcial-intelligence-is-changing-our-lives/we-need-a-cern-for-ai-in-europe (interviewing Professor Holger Hoos, a proponent of a CERN for AI); Laurie Clarke et al., *Rishi Sunak Wants to Lead the World on AI. The World ain't Listening*, POLITICO (June 5, 2023), https://www.politico.eu/article/rishi-sunak-ai-technology-wants-to-lead-the-world-on-ai-the-world-aint-listening/; Nicolas Mialhe, *Why We Need an Intergovernmental Panel for Artificial Intelligence*, UNITED NATIONS UNIVERSITY (Dec. 21, 2018), https://ourworld.unu.edu/en/why-we-need-an-intergovernmental-panel-for-artificial-intelligence.

commit to substantial financial investments in an autonomous institution. Exact emulation of the IPCC would also produce an inadequate institution—namely, one unequipped to keep pace with the rapid and evolving nature of AI risks.

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

1. CERN ANALYSIS

Cohesion, collaboration, 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.[124] 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 member states can call themselves CERN founding members. What's more, those eleven were relatively geographically proximate, held similar aspirations, and shared certain values and historical ties.

Today, geopolitical tensions have resulted in "the inability of countries to cooperate via international institutions to address policy problems that span borders."[125] Beyond the difficulties of forming an international entity posed by a competitive and hostile international climate, an AI initiative would face other barriers that did not apply to CERN. The initiative likely would need to involve many more countries as well as other types of actors, such as AI labs.[126] And, whereas CERN participants shared an understanding of the purpose of the organization and its potential outcomes, AI research raises more questions than answers—and those questions involve several politically and economically significant topics that, once answered, could change humankind.[127]

The impediments to an AI 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 the CERN Convention, which 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.[128]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.

Article II, Section 1 of the CERN Convention states that "[t]he Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character[.]"[129] Four things about that mission deserve attention: first, collaboration is mandated; second, the geographic scope is limited; third, the subject and purpose of research

---

[124] *See The History of CERN*, CERN, https://timeline.web.cern.ch/timeline-header/89 (last visited Sept. 17, 2023).
[125] ROBINSON, *supra* note 119.
[126] MARCUS & REUEL, *supra* note 87.
[127] *Id.*
[128] *See, e.g.*, Article II, § 1, CERN CONV.
[129] *Id.*

are made clear; and, fourth, a different part of the Article unequivocally prevents CERN from having "concern with work for military requirements[.]"[130]

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."[131] This direction evidenced a desire to "avoid duplicating research work[.]" This direction resulted in CERN becoming the focal point of a broad global network of researchers, thereby making CERN the indispensable center of international physics research.

Returning to an international AI research initiative, emulation of 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 organizations from quarrels over things like revenue-sharing agreements. On paper, it's easy to espouse a comparably simple mission for the initiative such as the identification of AI risks and analysis of means to mitigate those risks. In reality, the willingness of participants to 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 varied 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 responsibilities as well? Reuel and Marcus, 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's possible that an international AI 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 research initiative. Whereas the line from particle collisions to weaponization of that research is attenuated, national security observers have made clear that AI advances can and, according to some, must be integrated into a nation's defense strategy.[132] The same is true with respect to commercialization. Some countries, like the United Kingdom, have signaled a "pro-innovation" stance on AI development.[133] Such nations may not voluntarily share data, expertise, and other inputs that could assist with the initiative's aims.

Calls for a "CERN for AI" must address 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 research initiative.

---

[130] Article II, § 1, CERN CONV.
[131] Article II, § 7, CERN CONV.
[132] *See, e.g.*, Chuck Young, *How Artificial Intelligence Is Transforming National Security*, GAO (Apr. 19, 2022), https://www.gao.gov/blog/how-artificial-intelligence-transforming-national-security.
[133] *See, e.g.*, Astha Rajvanshi, *Rishi Sunak Wants the U.K. to Be a Key Player in Global AI Regulation*, TIME (June 14, 2023), https://time.com/6287253/uk-rishi-sunak-ai-regulation/.

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 made several significant voluntary actions to shore up the legitimacy and capacity of the organization. In particular, members committed to not only continue to pool their resources but also to automatically increase their annual financial contributions to CERN's budget.[134]

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 (1) the programs they participate in (i.e. the research they conduct and support) and (2) the size of their economy.[135] This mechanism anticipates and tolerates member states occasionally falling behind on their contributions—flexible payment plans allow such states to remain as CERN stakeholders and, in turn, facilitates the continued participation of a growing set of member states.[136] Members, though, have not exploited this flexibility. In fact, some states deduct their annual CERN dues from the foreign policy budget as a way to safeguard their contribution from political whims.[137] 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 be reinvested in that nation's economy.[138]

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."[139]

Another pivotal step in CERN's development and entrenchment as the "apex" of the European scientific community came when several members shuttered their respective national physics labs and directed those resources and personnel to CERN instead.[140] 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 lab. 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.[141] 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."[142]

---

[134] Dominique Pestre, SOME CHARACTERISTIC FEATURES OF CERN IN THE 1950S AND 1960S, STUDIES IN CERN HISTORY (1st. ed. 1989) at 3, cds.cern.ch/record/707442/files/CERN-CH-28.pdf?version=1.
[135] Article VII, CERN CONV.
[136] *See, e.g.*, Edwin Cartildge, *Greece May Not Be Able to Afford CERN*, SCIENCE (Apr. 3, 2012), https://www.science.org/content/article/greece-may-not-be-able-afford-cern (discussing flexible payment plans for Greece to remain affiliated with CERN through its austerity crisis).
[137] *See* ROBINSON, *supra* note 119, at 3 (internal citation omitted).
[138] *See* PESTRE, *supra* note 134, at 4.
[139] *Id.* at 3.
[140] KRIGE, *supra* note 18, at 16; PESTRE, *supra* note 134, at 6.
[141] ENGELEN & 'T HART, *supra* note 3, at 220.
[142] *Id.* at 218 (internal quotation omitted).

A similar level of resource consolidation seems unlikely in the AI context given the national security concerns and economic significance national leaders have placed on achieving AI superiority. South Korean leaders intend their country to become one of the three leading AI "powerhouses" by 2027.[143] President Emmanuel Macron of France has similar ambitions for his country. He recently pledged $500 million to spur the development of AI "champions."[144] President Biden has likewise stressed that he wants the U.S. to "lead the way toward responsible innovation."[145] That goal may foreclose substantive resource consolidation.[146] The Biden Administration, per Axios, has taken measures to deny China access to GPUs in order to advance the U.S.'s standing in the competition for "supremacy" in artificial intelligence.[147] 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.[148] 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.[149]

An international willingness to share resources and expertise for an AI initiative does not exist in the present geopolitical 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 on 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.[150] Case in point, U.S. law prohibits Japanese nationals from contributing to U.S.-sponsored sensitive scientific research.[151] Even if those security clearance issues were resolved, additional collaborative tasks, such as the exchange of sensitive information, would likely entail numerous

---

[143] Sheila Chiang, *South Korea wants to be a top A.I. hub - its memory chip dominance could be an advantage*, CNBC (July 5, 2023), https://www.cnbc.com/2023/07/06/south-koreas-dominance-in-memory-chips-an-advantage-in-ai-race.html.

[144] *Macron announces €500 million in funding for AI*, LE MONDE (June 14, 2023), https://www.lemonde.fr/en/france/article/2023/06/14/macron-wants-france-to-be-among-ai-leaders_6031624_7.html.

[145] Molly Nagle, *Biden White House, tech companies launch new safeguards around emerging AI technology*, ABC (July 21, 2023), https://abcnews.go.com/Politics/biden-white-house-tech-companies-launch-new-safeguards/story?id=101555314.

[146] Dave Lawler, *How the U.S. is trying to stay ahead of China in the AI race*, AXIOS (June 29, 2023), https://www.axios.com/2023/06/29/us-china-chip-export-restrictions-ai-race.

[147] *Id.*

[148] *Id.*

[149] *Id.*

[150] James L. Schoff, Douglas E. Rake & Joshua Levyal, *A High-Tech Alliance: Challenges and Opportunities for U.S.-Japan Science and Technology Collaboration*, CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE (July 29, 2021), https://carnegieendowment.org/2021/07/29/high-tech-alliance-challenges-and-opportunities-for-u.s.-japan-science-and-technology-collaboration-pub-85012.

[151] *Id.* More generally, foreign nationals often face substantial legal hurdles to contributing to national security projects. *See* Asha Balakrishnan, Susannah V. Howieson, & Elizabeth C. Lee, *Guidance for Employing Foreign Citizen Scientists and Engineers at Department of Defense Science and Technology Reinvention Laboratories*, SCIENCE & TECHNOLOGY INSTITUTE at 2 (2013), https://www.ida.org/-/media/feature/publications/g/gu/guidance-for-employing-foreign-citizen-scientists-and-engineers-at-department-of-defense-science-and/ida-d-4786.ashx (explaining the legal barriers that generally prevent foreign nationals from working with national security organizations).

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.[152]

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 research initiative than did 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 to more than a group of eleven states. Attempting to coordinate resource pooling from more actors as well as from more diverse actors will pose challenges. Yet, given the importance of such research being global, well-endowed, and 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[153]by "control[ing] CERN's activities in all matters, scientific, technical and administrative."[154] 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 IMF, that allocate voting power in proportion to financial contributions.[155] Scholars contend that the "one member, one vote" approach has created a "stable, level-playing field" among member states.[156] This theory is supported by the retention of this voting system through the expansion of the number of member states from 12 to 23—many of which contribute vastly different amounts in terms of financial resources and technical expertise.

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 non-unanimous decisions, the Council aims for unanimity, 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.[157] Whereas the Council tends to advance the intergovernmental and policy-driven aspects of CERN, the DG—informed by "elite scientists

[152] James L. Schoff et al., *supra* note 150.
[153] *Our Governance*, CERN, https://www.home.cern/about/who-we-are/our-governance (last accessed Sept. 20, 2023).
[154] ENGELEN & 'T HART, *supra* note 4, at 221.
[155] *IMF Members' Quotas and Voting Power, and IMF Board of Governors*, INTERNATIONAL MONETARY FUND, https://www.imf.org/en/About/executive-board/members-quotas (last accessed Sept. 21, 2023).
[156] ENGELEN & 'T HART, *supra* note 4, at 221-22.
[157] Welcome to the CERN Council, CERN (Sept. 21, 2023), https://council.web.cern.ch/en.

whispering in [their] ear"—compliments the Council by focusing on transnational and scientific priorities.[158]

The Council appoints the DG, who usually serves for a single five-year term.[159] 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."[160] 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.[161] The combination of a single-term and a built-in transition period diverged from the typical approach taken by similar organizations and may have played a role in CERN experiencing relatively more stability than similarly oriented organizations.[162] A single term may reduce the odds of the DG steering CERN's work in light of temporary political conditions.[163] Transition periods may prevent a sudden loss of institutional knowledge that can hinder an organization's efficacy.[164] Cumulatively, these structural decisions have likely contributed to observers citing CERN as a model for other organizations to follow.[165]

CERN's governance systems include other unique features worthy of potential emulation. Several advisory bodies within CERN play a role in its decision making, 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 act as incubators for the next generation of CERN leaders. CERN also hosts several "user committees," which develop as needed around the operation of specific machines. These committees in turn send representatives to the Advisory Committee of CERN Users, which more generally advocates on behalf of CERN users.[166] 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."[167]

---

[158] ENGELEN & 'T HART, *supra* note 4, at 217.
[159] *Cf.* PRESS RELEASE, CERN, *CERN Council appoints Fabiola Gianotti for second term of office as CERN Director General* (Nov. 6, 2019), https://home.cern/news/press-release/cern/cern-council-appoints-fabiola-gianotti-second-term-office-cern-director.
[160] PESTRE, *supra* note 134, at 3.
[161] ROBINSON, *supra* note 119, at 8.
[162] *Cf.* Rolf Heuer, *Science for Good: Governance is the Key*, CERN (Jan. 23, 2015), https://home.cern/news/opinion/cern/science-good-governance-key (asserting that CERN's governance structure has fostered stability).
[163] *Cf.* Tom Ginsburg, James Melton & Zachary Elkins, *On the Evasion of Executive Term Limits*, UNIVERSITY OF CHICAGO PUBLIC LAW & LEGAL THEORY WORKING PAPER NO. 328 at 5 (2010) (examining the merits of a single-term approach in presidential systems and observing that the resulting insulation "can facilitate the undertaking of policies that might entail short term costs, but produce benefits in the mid-term.").
[164] *Cf.* John D. Gregory, *The Director's Creed: Give, Get or Get Off*, THE PHILANTHROPIST JOURNAL (Mar. 1, 1985), https://thephilanthropist.ca/1985/03/the-directors-creed-give-get-or-get-off/ (discussing various strategies to reduce instability amid a change in executive leadership).
[165] *See* Katharina E. Hone & Jovan Kurbalija, *Accelerating Basic Science in an Intergovernmental Framework: Learning from CERN's Science Diplomacy*, 9 GLOBAL POLICY 67, 68-69 (2018), https://onlinelibrary.wiley.com/doi/pdf/10.1111/1758-5899.12589 (remarking on CERN's governance structure as a potential model for other organizations).
[166] Anaïs. Schaeffer, *ACCU: A committee that addresses users' needs*, CERN (June 4, 2013), https://home.cern/news/news/cern/accu-committee-addresses-users-needs.
[167] ENGELEN & 'T HART, *supra* note 4, at 220.

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.[168]

Replication or even partial emulation of such a governance structure by a "CERN for AI" is a tall task where success 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 "pure" research mission in place and the identification of specific and well-defined experiments to realize that mission, however, it may become easier for governments, labs, and other AI stakeholders to grant the international AI 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' governments. 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 who, 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.

A counterfactual that reinforces CERN as an exemplary model

Prior to discussing lessons an international AI research initiative could and should learn from the IPCC, it's worth looking to the U.S.'s struggles to undertake world-leading particle physics around the same time as CERN was developing as a sort of counterfactual that reinforces the merits of studying and emulating CERN.

The importance of scale—with respect to funds and expertise—and a culture of collaboration to CERN's work helps to explain why similar, albeit more insular particle physics labs never got past the ideation phase. A comparison of efforts to establish and sustain "big science experiments" in the US make the distinct factors that gave rise to CERN all the more clear. This section contains just a few snapshots of the barriers to US efforts to keep pace with CERN, though a fuller examination of such barriers likely merits attention.

By 1976, the US, USSR, and a smattering of other countries dreamed up an effort to build a machine larger than anything at CERN.[169] The superpowers, though, were the main players (and planned investors). Following discussion of a "world machine" at international meetings, the duo tapped the International Union of Pure and Applied Physics to study the idea further and "take the organizational initiative."[170] At the same time, both the US and USSR were

[168] *Id.* at 230-31.
[169] Walter Sullivan, *Physicists Hoping to Build 30-Mile Atom Device to Explore Matter*, THE N.Y. TIMES (Oct. 10, 1976), https://www.nytimes.com/1976/10/10/archives/physicists-hoping-to-build-30mile-atom-device-to-explore-matter.html?searchResultPosition=24.
[170] *Id.*

discussing expansions of their own national labs—the BNL and UNK, respectively. This "world machine," of course, did not come about.[171]

Absent progress on the "world machine," the US tried in the early 1980s to move ahead with updates to the Brookhaven National Laboratory (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 required (1) too much money, (2) would require much more, (3) time that the US didn't have if it wanted to stay ahead of CERN, and, (4) political will that scientists could not muster.[172] The *Times* called for the U.S. to consider increasing its participation and investment in CERN[173]—just as member states ditched their national labs early in CERN's development.

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. To make matters worse, U.S. labs that did not have CERN's relatively stable financial situation had to compete with concerns about a smaller economic pie. As detailed by Robert Crease and Charles Mann, the US 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."[174]

By the late 1980s, looking back at decades of failed attempts to get the U.S.'s on the level of CERN, the *Times* argued that "America has no coherent policy for supporting scientific research."[175] By way of example, the *Times* examined Congress's consideration of a proposed "superconducting super collider," a $5 billion machine for particle physics experiments.[176] Congress, per the *Times*, had not assessed the trade-offs that the "SSC" would require.[177] The proposal seemed to ignore other obvious options to achieve similar research—namely, greater participation in CERN. In short, the Times chalked the SSC up to an idea that may have been justified in "different times."[178]

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.[179] Representatives likely took issue with the fact that CERN was working on a similar collider at nearly one-eighth the cost of the SSC.[180] 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."[181] CERN

---

[171] *Id.*
[172] *The Trouble With Isabelle*, THE N.Y. TIMES (June 24, 1982), https://www.nytimes.com/1982/06/24/opinion/the-trouble-with-isabelle.html?searchResultPosition=35.
[173] *Id.*
[174] Robert Crease & Charles Mann, Opinion, *Gambling With The Future of Physics*, THE N.Y. TIMES (Dec. 5, 1982), https:/www.nytimes.com/1982/12/05/magazine/gambling-with-the-future-of-physics.html?searchResultPosition=40.
[175] Editorial, *The Unaffordable Atom Smasher*, THE N.Y. TIMES (Nov. 16, 1988), https://www.nytimes.com/1988/11/16/opinion/the-unaffordable-atom-smasher.html?searchResultPosition=85.
[176] *Id.*
[177] *Id.*
[178] *Id.*
[179] Barry James, *Europe Is Ready to Pick Up the Pieces in Particle Research*, THE N.Y. TIMES (June 20, 1992), https://www.nytimes.com/1992/06/20/IHT-europe-is-ready-to-pick-up-the-pieces-in-particle-research.html?searchResultPosition=98.
[180] *Id.*
[181] *Id.*

researchers also commented that legislators may have recognized that Europe had successfully executed a trans-Atlantic brain drain that meant more of the U.S.'s top physicists did research in Europe than vice versa.[182] Reversing that trend, of course, would add to the costs and logistics of getting the SSC going.

This brief examination of the U.S.'s struggles to go-it-alone on "big" science reinforces that such endeavors require scale in more ways than one. In particular, the U.S.'s shortcomings demonstrate that financial might alone cannot guarantee the success of such initiatives. CERN's founding member states entered into a pact of mutually-assured collaboration in which every member had a stake in keeping the organization going, regardless of changes in their domestic politics. Absent that sort of peer pressure, a single nation may not have the necessary defenses to push back when new national priorities arise and command more resources and political attention. And, though directly copying CERN in the AI context is likely not possible nor advisable given key differences in the two projects, proponents of safe AI development would be wise to incorporate the collaborative ethos that shapes each aspect of CERN.

Lessons from CERN for an international AI research initiative

Though this section has interspersed CERN lessons for an AI initiative throughout its analysis, it's worth consolidating and clarifying those lessons here.

Lesson #1: Independent, "pure" risk research is not by accident

CERN has myriad structural, cultural, and legal aspects in place to facilitate "pure" research. This deliberate and comprehensive approach to directing research toward whatever the science should be emulated by the AI initiative. Absent such a "layered" approach to safeguarding research from political aims and commercial opportunities, a lab with such expertise and resources may struggle to remain true to the institution's goals. Moreover, even when it appears as though an organization has sufficiently taken steps to protect the autonomy of researchers, a glance back at CERN's history makes clear that actors will inevitably find cause to test those efforts.

It follows that the AI initiative should copy and expand on the wise decisions made by CERN's founders. One particularly important step may be to ingrain a "pure" research focus into the initiative's DNA by, 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. Attempts by the U.S. to develop a similarly large and ambitious project floundered because it lacked partners that could fill any financial gaps upon tough economic times or challenging political conditions. So, whereas political opposition to the U.S. efforts derailed those efforts, CERN has continued to evolve and expand despite members occasionally having to forgo or limit their annual contribution.

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 is resilient. Of

[182] *Id.*

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 superiority of doing particle research at scale both in terms of 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 brightest in the field.

Gathering AI experts at a single location 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 across America 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. I welcome and encourage more people to expand on this list. In the same way that CERN's success resulted from creative institutional thinkers implementing novel governance and operational mechanisms, the difficulties posed by launching an international AI research initiative will likely not be resolved by looking at tired and outdated concepts.

2. IPCC ANALYSIS

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 the initiative are not willing to lend their financial resources and technical expertise to an AI research hub modeled after CERN, then the IPCC's model warrants especially close attention by advocates for an international AI research initiative. This alternative would not produce the same positive feedback loop 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.

Regardless of what form the initiative takes—closer to CERN, the IPCC, or a hybrid—study of the IPCC's mission, network of experts, and work product can inform how best to go about identifying and prioritizing AI risks

Overview of the IPCC

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."[183] The drafting stage involves the formation of "author teams" tasked with producing a topic-specific portion of the report. Teams are assigned to one of three working groups: The Physical Science Basis of Climate Change; Climate Change Impacts, Adaptation and Vulnerability; and Mitigation of Climate Change. A Technical Support Unit of IPCC provides staff support to each working group.

The author selection process proceeds in two steps. First, governments and IPCC observer organizations nominate authors based on their submission of detailed CVs.[184] In the lead up to the Fifth Assessment Report, published in 2014, the IPCC received 3,598 such nominations.[185] More recently, about 2,800 experts were nominated to author the Sixth Assessment Report.[186] Second, the IPCC selects the authors based on their expertise and in furtherance 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, individuals with industry experience, and individuals from non-profit organizations.[187] Inclusivity, though, does not foreclose a competitive process. The Fifth Assessment Report, for example, included 831 authors, or only one quarter of the nominees.[188]

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."[189] Throughout the drafting process, the authors try to produce a "[b]alanced assessment of the full range of scientific views" and 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.[190]

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.[191] Reliance on experts, though, does not mean this process is not open and participatory. The IPCC 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."[192] This includes scholars, industry representatives, nonprofit experts, and others.[193]

---

[183] *IPCC Factsheet: How does the IPCC review process work?*, INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE 1 (Jan. 15, 2015), https://www.ipcc.ch/site/assets/uploads/2018/02/FS_review_process.pdf [hereinafter, *Review Process*].
[184] *Id.* at 1.
[185] *Id.* at 2.
[186] *Id.*
[187] *Id.*
[188] *Id.*
[189] *Id.* at 1.
[190] *IPCC, How does the IPCC select its authors?*, 2 https://www.ipcc.ch/site/assets/uploads/2021/07/AR6_FS_select.pdf (last accessed Jan. 24, 2024),
[191] *Id.* at 2.
[192] *Id.* at 1.
[193] Nick Stockton, *The US Won't Pay For the World's Best Climate Science*, WIRED (Aug. 11, 2017), https://www.wired.com/story/the-us-wont-pay-for-the-worlds-best-climate-science/

Individuals who want to comment can simply go through a self-declaration process to establish their credentials as an expert.[194]

The 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."[195] These revisions demonstrate that the initial round of review comments are not for show nor are the preliminary stages intended only for highly technical audiences. Once the Second Drafts and SPM are ready, the IPCC invites the first round experts to comment once more, invites other experts to do the same, and provides governments with a chance to review and comment on the drafts. These, again, are not empty gestures. A variety of stakeholders take advantage of these input opportunities, illustrated again by engagement with Working Group I's section of the assessment: exactly 800 experts and twenty-six governments commented on the Second Order Draft.[196]

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.[197]

All of this work is at the direction and mandate of the IPCC Panel, a body made up of 195 member governments.[198] 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[.]"[199]

The IPCC's funding structure, however, provides a cautionary rather than exemplary tale. About 25 countries voluntarily fund the IPCC's "shoestring operations."[200] This limited donor base 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 US support, depriving the IPCC of $2 million, or nearly half of the IPCC's annual operating expenses.[201]

Lessons from the IPCC for an international AI research initiative

Prior to exploring the specific lessons that proponents of an international AI research initiative could glean from the IPCC, it is important to first establish the extent to which the IPCC serves as a valid comparison. Several similarities between the AI risk research and climate research lend support for establishing an IPCC for AI. First, both the climate and AI can be characterized as global systems.[202] That aspect of the climate underpinned the IPCC's work to

---

[194] IPCC, *supra* note 190, at 1.
[195] *Id.*
[196] *Id.* at 2.
[197] *Id*. at 2.
[198] *Structure of the IPCC*, IPCC, https://www.ipcc.ch/about/structure/ (last accessed Sept. 21, 2023).
[199] *Id.*
[200] STOCKTON, *supra* note 193.
[201] *Id.*
[202] Pablo Hernandez-Lagos, *The Beijing dilemma: Dependencies in global artificial intelligence research*, BROOKINGS (May 7, 2024), https://www.brookings.edu/articles/the-beijing-dilemma-dependencies-in-global-

become a "globally-representative scientific institution[.]" For reasons discussed throughout this paper, AI research would similarly benefit from including an international set of experts with a range of perspectives.

Second, in the same way economic, political, and cultural differences have thwarted the formation of a global climate institution with a broad regulatory mandate,[203] analogous differences exist among AI stakeholders, as mentioned above. 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.[204] A similar desire exists in the AI context.[205]

Importantly, 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.[206]

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."[207] 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.[208] The role of IPCC assessments in guiding international action on climate, however, suggests that the inclusion of national governments is a net positive. This argument points to evidence that the involvement of national governments "increase[s] the political salience of climate change through linking knowledge production with national governments and the UN system."

Finally, despite the high demand for climate research, member nations have a limited willingness to fund it. 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.[209] If nations are similarly stingy with respect to funding an international AI research initiative, then the

---

artificial-intelligence-research/ ("The reality is that even advanced countries must tap into knowledge and talent from wherever talent and knowledge are available if they are to lead the advance of the field."); *see* Emma Klein & Stewart Patrick, *Envisioning a Global Regime Complex to Govern Artificial Intelligence*, CARNEGIE (Mar. 21, 2024), https://carnegieendowment.org/research/2024/03/envisioning-a-global-regime-complex-to-govern-artificial-intelligence?lang=en ("[I]nternational governance will also be critical, given the global reach and ramifications of AI."); *cf.* Richard Vanderford, *Risk Professionals Take On Emergent AI, Climate Change and Global Conflict*, WALL STREET JOURNAL (May 9, 2024), https://www.wsj.com/articles/risk-professionals-take-on-emergent-ai-climate-change-and-global-conflict-a5197423 (collecting advice from risk researchers on how to respond to AI, climate change, and general political instability).

[203] G.A. RES. 44/207, RESOLUTION, PROTECTION OF THE GLOBAL CLIMATE FOR PRESENT AND FUTURE GENERATIONS OF MANKIND (Dec. 22, 1989) (endorsing the creation of the IPCC and tasking it with developing an overview of the state of knowledge of the science of climate change).

[204] *Id.*

[205] *See, e.g.*, Chris Vallance, *AI risks destabilising world, deputy PM to tell UN*, BBC (Sept. 21, 2023), https://www.bbc.com/news/technology-66879709.

[206] PAGLIA & PARKER, *supra* note 10, at 299.

[207] *Id.*

[208] *Id.* at 300.

[209] STOCKTON, *supra* note 193.

barebones tactics and processes employed by the IPCC will provide even more guidance for the creation of such an initiative. The current trend of nations identifying their own costly regulatory regimes[210] suggests that research funds may indeed be limited.

Limits to learning from the IPCC

Other aspects of the IPCC's work and structure cut against its value as a model institution for an international AI research initiative. Most glaringly, the IPCC's reliance on voluntary participation by experts may only be possible because of the lengthy time between each of its assessments, typically six to seven years.[211] A much shorter period might cause fewer experts to volunteer their time, regardless of the prestige that comes with contribution to an assessment.[212] Interviews of IPCC experts confirms they find the work "intense, stressful, and unsustainable."[213] Yet, an international AI 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.[214]

While the IPCC relied on an unrepresentative set of experts to produce its first assessment in 1960,[215] the international community nevertheless 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."[216] About 100 authors contributed to that first assessment, but fewer than 20 authors represented the Global South and fewer than 10 were women.[217]

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 importance. Regardless of the rationale, the international AI 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

---

[210] ECONOMIST INTEL., *How governments are looking to regulate AI* (July 21, 2023), https://www.eiu.com/n/how-governments-are-looking-to-regulate-ai/.
[211] Press Release, IPCC, *IPCC meets to approve the final component of the Sixth Assessment Report* (Mar. 13, 2023) (on file with author).
[212] See Allister Doyle, *'Ridiculous' length? How to make IPCC climate science reports an easier read*, CONTEXT (Apr. 6, 2022), https://www.context.news/climate-risks/how-to-make-ipcc-climate-science-reports-an-easier-read (pointing out the prestige associated with volunteering).
[213] Ayesha Tandon, *Analysis: How the diversity of the IPCC authors has changed over three decades*, CARBON BRIEF (Mar. 15, 2023, 7:00 AM), https://www.carbonbrief.org/analysis-how-the-diversity-of-ipcc-authors-has-changed-over-three-decades/.
[214] *See, e.g.*, Will Henshall, *4 Charts That Show Why AI Progress Is Unlikely to Slow Down*, TIME (Aug. 2, 2023, 4:50 PM), https://time.com/6300942/ai-progress-charts/; Shana Lynch, *AI Benchmarks Hit Saturation*, STAN. HUM.-CENTERED A.I. (Apr. 3, 2023), https://hai.stanford.edu/news/ai-benchmarks-hit-saturation.
[215] TANDON, *supra* note 194.
[216] PAGLIA & PARKER, *supra* note 10, at 296.
[217] TANDON, *supra* note 213.

North,[218] inadequate engagement with Global South values, perspectives, and experts would minimize the impact of any research overseen 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 community of AI experts who could potentially join an AI research initiative. For instance, efforts are underway to bolster the number of students pursuing AI-related graduate degrees in Africa to bolster the continent's expertise in the area.[219] Similarly, AI stakeholders in Latin and South America fear that a lack of AI expertise may partially explain why their representatives have had little sway in AI governance talks.[220] 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.[221]

It follows that an international AI research initiative may have to develop more intentional processes to identify a global set of AI experts. The IPCC's approach, though increasingly inclusive in recent years, certainly does not guarantee a representative set of assessment authors. Still, only around 30 percent of the authors of the latest IPCC assessment were women and about 40 percent represented the Global South.[222] To improve on these numbers, the initiative should not only consider a quota system for its experts but also actively engage in efforts to foster more AI expertise around the world, perhaps by using some of its funds to invest in regional AI research centers.

Finally, an international AI research initiative will likely require a different funding system than the IPCC. At minimum, the initiative may need a larger budget for at least two reasons: first, the scarcity of AI experts may militate against a general willingness to volunteer for the initiative; second, verification and summarization of AI research may require more capital investments than the same task when done in the climate science context.

More importantly, the initiative needs a more reliable source of funds. Though the IPCC managed to circumvent the funding shortfall induced by the Trump Administration, it did so by placing a greater financial burden on other countries and by tapping into limited financial reserves.[223] If the initiative is going to reliably produce consensus reports on the risks posed by the latest AI models, then it will need to make investments in increasing its team of AI experts and in growing the technical infrastructure necessary to analyze the latest AI research.

With those similarities and differences made clear, advocates for an IPCC for AI should specifically aim to emulate the following lessons:

---

[218] *See, e.g.*, Robert Muggah & Ilona Szabó, *Artificial Intelligence Will Entrench Global Inequality*, FOREIGN POLICY MAGAZINE (May 29, 2023, 8:00 AM), https://foreignpolicy.com/2023/05/29/ai-regulation-global-south-artificial-intelligence/.

[219] *See, e.g.*, Eve Ruwoko, *5,000 PhD scholars to meet Africa's growing AI needs*, UNIV. WORLD NEWS (Feb. 1, 2022), https://www.universityworldnews.com/post.php?story=20220130114337493.

[220] *The global AI agenda: Latin America*, MIT TECH. REV., (June 8, 2020), https://www.technologyreview.com/2020/06/08/1002864/the-global-ai-agenda-latin-america/.

[221] *See, e.g.*, Karen Hao, *China's path to AI domination has a problem: brain drain*, MIT TECH. REV., (Aug. 7, 2019), https://www.technologyreview.com/2019/08/07/133830/china-ai-domination-losing-talent-to-us/.

[222] TANDON, *supra* note 213.

[223] Brenda Ekwurzel, *Donald Trump Ends IPCC Funding and 'Abandons Global Science Leadership*', ECOLOGIST (Aug. 17, 2017), https://theecologist.org/2017/aug/17/donald-trump-ends-ipcc-funding-and-abandons-global-science-leadership.

Lesson #1: Consensus can drive action

The IPCC review process is akin to a "red-teaming" exercise in which every conclusion must survive analysis by nations inclined to skeptically view climate science.[224] 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 scientific integrity. The value of that integrity cannot be understated. Eric Paglia and Charles Parker 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."[225]

Absent a consensus-based approach, concerns about the IPCC struggles with achieving the proper balance of perspectives and expertise might completely undermine its effectiveness, rather than slightly hinder it. More specifically, the IPCC's thorough and participatory drafting process remedies 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[.]"[226]

The 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 AI ethics as well as those most concerned about AI safety should have their work reviewed by the initiative and be engaged in the production of its reports.[227] If the initiative appears predisposed—by personnel, process, or both—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 abstain from acting on the initiative's conclusions. Of course, achieving consensus often comes at a cost. 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.

Lesson #2: Informed action at any level is beneficial

Authors draft IPCC reports in an accessible and transparent fashion. This democratization of complex and esoteric research from around the world makes it easier for stakeholders to take responsive action. Political leaders from city councilors to prime ministers "deeply value[]" IPCC assessments.[228] The same is true of members of the "business community, civil society, advocacy organizations and other stakeholders."[229] Assessments even serve as a motivator for more demanding activists, such as Greta Thunberg.[230] The steps taken by these actors to reduce climate change may, in isolation, seem insignificant; yet, climate action by smaller actors like

---

[224] John Holdren, *The Perversity of the Climate Science Kangaroo Court*, THE BOSTON GLOBE (July 15, 2017), https://www.bostonglobe.com/opinion/2017/07/24/the-perversity-red-teaming-climate-science/VkT05883ajZaTPMbrP3wpJ/story.html; *see* Jessica Ji, *What Does AI Red-Teaming Actually Mean?*, CSET (Oct. 24, 2023), https://cset.georgetown.edu/article/what-does-ai-red-teaming-actually-mean/ (analyzing red-teaming in the context of AI research wherein red team is tasked with looking for faults in a proposal.).
[225] PAGLIA & PARKER, *supra* note 10, at 296.
[226] *Id.* at 297.
[227] Kif Leswing, Parrots, *Paper Clips and Safety vs. Ethics: Why the Artificial Intelligence Debate Sounds like a Foreign Language*, CNBC (May 21, 2023), https://www.cnbc.com/2023/05/21/ai-terminology-explained-paper-clips-parrots-and-safety-vs-ethics.html.
[228] PAGLIA & PARKER, *supra* note 10, at 296.
[229] *Id.*
[230] *Id.*

city governments has been touted as an important and effective way of steering the world away from a more dire future.[231]

In a related fashion, the current disparity in AI expertise and resources makes the democratization of information about AI risks all the more important. Not every country has to take responsive action based on those risks to help reduce the overall likelihood of AI risks. Case in point, one AI risk scenario involves totalitarian regimes using the technology to expand their control over civil society.[232] If reports by the international AI research initiative assisted even a few countries predisposed to such regimes enact responsive regulations, then the overall risks posed by AI would fall.

The case for an IPCC for AI

The IPCC model does not map perfectly onto the AI regulatory environment. Still, even if slightly adjusted to accommodate the novel challenges associated with regulating AI, such a model still carries tremendous potential to reduce AI risks. The upshot of the IPCC's procedures and work products is that quality, trustworthy information can stimulate political negotiations and policy innovations.[233] Additionally, the IPCC makes clear that aggregating and sorting research is a social good that can be done relatively cheaply and with a barebones operational footprint.

Though many stakeholders concerned about AI risks might prefer to focus financial and political capital on creating a large international organization with a broad regulatory mandate, the success of the IPCC suggests that a leaner, research-driven organization may develop in a faster fashion and facilitate homegrown regulatory solutions that nevertheless have global benefits.

Specific Structural Recommendations for an International AI Research Initiative from CERN and the IPCC

The unique risk profile posed by advances in AI likely requires the creation of a bespoke international institution. Advances in AI and, by extension, new risks from AI develop at too rapid a pace to justify emulation of the IPCC's approach. CERN's reliance on a small group of countries with shared concerns disqualifies it as a perfect fit for the global research envisioned for an international AI risk initiative. That said, the more high-level institutional features that have sustained the important work of CERN and the IPCC do merit close study and some degree of replication by AI risk researchers. Any international AI risk research initiative should embody the most crucial attributes for success gleaned from the prior study of those institutions: clarity of mission, sustainable funding, expert recruitment, and robust governance.

---

[231] *Cities: a 'Cause of and Solution to' Climate Change*, UN NEWS (Sept. 18, 2019), https://news.un.org/en/story/2019/09/1046662.

[232] *See, e.g.*, Simon McCarthy-Jones, *Artificial Intelligence is a Totalitarian's Dream - Here's How to Take Power Back*, THE CONVERSATION (Aug. 12, 2020), https://theconversation.com/artificial-intelligence-is-a-totalitarians-dream-heres-how-to-take-power-back-143722.

[233] PAGLIA & PARKER, *supra* note 10, at 301.

Clarity of Mission

Both CERN and the IPCC have withstood the ever-present temptation to drift from their respective original purposes thanks to each institution having a clear, focused mission. For an AI research initiative, this mission should center on understanding and mitigating long-term risks posed by AI technologies globally, with an emphasis on investigating issues that may not be prioritized by private actors. This includes the unintended consequences of AI deployment that could affect social, economic, and ethical aspects of life worldwide, particularly in regions that are often overlooked like the Global South.[234] The initiative should aim to provide a comprehensive, impartial assessment of AI risks, akin to how CERN delves into fundamental particle physics without military applications.

This clarity will guide all other aspects of the initiative, from funding to research priorities. A specific mission that forecloses certain applications of its research will also increase the likelihood of global participation in the initiative. An alternative arrangement in which researchers could use their findings to pursue private, commercial ends would likely skew the research agenda and diminish the rate of participation by experts from nations less likely to benefit from those commercial endeavors.

Creation of a Sustainable Funding System

Following the model of CERN, where member states contribute based on economic size and participation levels, the AI initiative should establish a funding mechanism that is both fair and adaptable to the economic realities of its members. This system must be designed to withstand geopolitical and economic fluctuations, ensuring continuous and stable financial support. Lessons from the IPCC, which operates on a tighter budget with contributions from a broader base of countries, also suggest the value of a diverse funding model that can leverage smaller contributions from a larger pool of countries to sustain operations.

One particular funding approach warrants additional study: a research initiative could incorporate a principle where countries that contribute more significantly to the risks associated with AI, either through intensive development or deployment practices, are required to contribute more to the initiative's budget. This approach not only aligns financial responsibility with the source of risk but also incentivizes nations to more responsibly manage their AI development practices. By scaling contributions according to the level of risk each country introduces into the global system, the initiative can ensure that those benefiting most from AI advancements also invest appropriately in mitigating any negative consequences. This funding model fosters a sense of fairness and shared responsibility, which is crucial for securing broad international support and commitment.

[234] *Cf.* Chinasa T. Okolo, *AI in the Global South: Opportunities and challenges towards more inclusive governance*, BROOKINGS (Nov. 1, 2023), https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance/ (highlighting that AI may generate progress on important problems across the Global South but flagging that the spread of AI also poses new, distinct challenges for these regions).

Recruitment and Retention of Foremost AI Experts

To forge a path similar to CERN's success in attracting top-tier physicists, the AI initiative should focus on creating an attractive environment for the world's leading AI researchers. This includes competitive compensation, resources for cutting-edge research, and a collaborative atmosphere that transcends national borders. Moreover, the initiative should actively engage with emerging AI hubs around the world to ensure a diverse and inclusive representation of global talent, which enriches the research and increases the initiative's relevance across different geopolitical landscapes. Akin to the IPCC, the research initiative should regularly share its results in a practical, understandable format with regulators and stakeholders. Researchers who contribute to such publications will rightfully receive the sort of recognition that will make them more likely to persist with risk research rather than to jump for private research opportunities.

The initiative will have to cultivate a diverse and interdisciplinary group of researchers, which may benefit from creating specific talent pipelines to the initiative from foremost academic institutions around the world. By way of example, if the top computer science students from specific programs were automatically offered jobs with the initiative, then it may be the case that the initiative will quickly be regarded as the place to be for researchers aiming to learn from and work with preeminent scholars. Similar pipelines could be developed for a range of disciplines.

Governance Systems that Insulate Research from Political Shifts

Governance of an initiative should protect the integrity of research against the sway of political changes, drawing on CERN's model of scientific autonomy. Similar to the IPCC, the initiative should also feature a flat hierarchy where decisions are driven by consensus among scientists rather than imposed by political agendas. However, recognizing the dynamic and evolving nature of AI technology, the governance structure should also be adaptable, enabling swift responses to new scientific findings and technological advancements. This flexibility will be crucial in maintaining the relevance and effectiveness of the initiative as it navigates the rapidly changing field of AI.

Moving Forward with Implementation

CERN and the IPCC did not become world-renowned institutions by mistake. Their respective success reflects unique governing mechanisms tailored to their missions. The proposed international AI risk research initiative must be a collaborative effort, involving not only nations but also private sectors, academia, and non-governmental organizations. This multi-stakeholder approach will enrich the initiative's perspectives and capabilities, facilitating a comprehensive understanding of AI risks and the development of robust mitigation strategies.

By integrating these attributes—clarity of mission, sustainable funding, expert recruitment, and insulated governance—the initiative can serve as a beacon of global cooperation and scientific excellence in AI research. This model would not only address the unique challenges posed by AI but also set a precedent for more international collaboration in other emerging technological fields, such as quantum computing.

CONCLUSION

Enacting regulation that lacks grounding in the latest research is akin to sailing without a compass—you'll move but likely not in the right direction. A virtuous research-regulation cycle prevents hasty and harmful regulation from being enacted or from persisting for too long. This paper makes the case for more attention being spent on the development of an international AI 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, scope, 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, permit some of its finite resources be used for commercial purposes, and produce research unlikely to be accepted by certain countries. An international AI research initiative, in contrast, could mirror CERN or the IPCC by welcoming a global set of AI experts who, collectively, can conduct "pure" AI risk research that is more likely to be accepted and acted on by countries, AI labs, and other stakeholders.

CERN and the IPCC stand out as modes 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 efforts in which a global set of AI experts analyze the latest AI research and, with guidance from national governments, produce consensus reports on the most pressing AI risks and the best steps to reduce those risks.

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 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—investments in pure AI risk research conducted by an international initiative will prevent regulators from becoming reliant upon the labs themselves for information on AI risks and will ensure consideration of AI's risks from a global perspective.