CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

MARCH 2024

Envisioning a Global Regime Complex to Govern Artificial Intelligence

Emma Klein and Stewart Patrick

Envisioning a Global Regime Complex to Govern Artificial Intelligence

Emma Klein and Stewart Patrick

© 2024 Carnegie Endowment for International Peace. All rights reserved.

Carnegie does not take institutional positions on public policy issues; the views represented herein are those of the author(s) and do not necessarily reflect the views of Carnegie, its staff, or its trustees.

No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from the Carnegie Endowment for International Peace. Please direct inquiries to:

Carnegie Endowment for International Peace
Publications Department
1779 Massachusetts Avenue NW
Washington, DC 20036
P: + 1 202 483 7600
F: + 1 202 483 1840
CarnegieEndowment.org

This publication can be downloaded at no cost at CarnegieEndowment.org.

Contents

Abbreviations

AI — artificial intelligence

CBM — confidence-building measure

CERN — European Organization for Nuclear Research

EU — European Union

FATF — Financial Action Task Force

HLAB — United Nations High-Level Advisory Body on Artificial Intelligence

IAEA — International Atomic Energy Agency

ICAO — International Civil Aviation Organization

ILO — International Labour Organization

IMF — International Monetary Fund

IMO — International Maritime Organization

IP — internet protocol

IPCC — Intergovernmental Panel on Climate Change

LAWS — lethal autonomous weapons systems

LMIC — low- and middle-income country

NPT — Treaty on the Non-Proliferation of Nuclear Weapons

OECD — Organisation for Economic Co-operation and Development

SDG — sustainable development goal

UN — United Nations

UNEP — United Nations Environment Programme

UNESCO — United Nations Educational, Scientific and Cultural Organization

UNICEF — United Nations Children's Fund

WHO — World Health Organization

WMO — World Meteorological Organization

Introduction

In spring 2023, OpenAI cofounders Sam Altman, Greg Brockman, and Ilya Sutskever proposed an "IAEA [International Atomic Energy Agency] for superintelligence efforts" to govern high-capability systems, noting the potential genesis of superintelligence in rapidly advancing artificial intelligence (AI) models.[1] Soon after, United Nations (UN) Secretary-General António Guterres lent his support to this idea.[2] In the following months, others suggested an array of global institutions that could be created to regulate this technology, based on models such as the Intergovernmental Panel on Climate Change (IPCC) and the International Civil Aviation Organization (ICAO).[3]

Less than a year later, the search for a single institutional solution has faded. The challenges that AI presents are too multifaceted, the relevant actors too varied, and the geopolitical situation too complicated for any one global body to tackle by itself. Instead, many expect the emergence of overlapping institutions designed to advance and govern specific uses and impacts of AI.[4] Illustrative of this shift in thinking is the preliminary report of the UN High-Level Advisory Body on AI (HLAB), produced by thirty-nine expert members and released in December 2023.[5] Although the report does not address how closely linked global AI governance institutions should be—and whether there should be "individual institutions" or a "network of institutions"—it presumes there will be multiple. The ultimate content and contours of this governance arrangement will reflect several functional imperatives, as refracted through the interests, values, and capabilities of powerful public and private actors with a stake in AI's future.

As this alternative approach to AI governance gains traction in policy discussions, several analytical gaps remain. First, few analysts have explicitly framed the anticipated governance framework for AI as a regime complex, much less grappled with the implications of such a complicated institutional design. Second, while various stakeholders, including the HLAB, have enumerated a welter of institutional analogies from other fields for particular governance functions, they have seldom interrogated the relevance of these models in any detail.[6] Third, few experts have explored how geopolitical dynamics, including strategic rivalry, will shape, and likely constrain, the creation of international institutions to govern AI. This working paper aims to fill these gaps and alert policymakers to the possibilities, dilemmas, and trade-offs that may arise as they design—and ideally seek to reconcile—multiple governance arrangements.

Global AI governance will inevitably involve some fragmentation. The history of internet governance, including debates over the appropriate regulatory role of governments, has illuminated the distinct orientations of the United States, the European Union (EU), and China toward the global digital order, characterized as "market-driven," "rights-driven," and "state-driven" models, respectively.[7] While the United States has championed a limited government role over the internet and deferred to private technology companies so as to support freedom of speech and technological innovation, the EU has pursued a greater regulatory role to protect other human rights, including privacy, and China has assumed complete state control, with extensive censorship and surveillance capabilities.

These differences—and the challenge of reaching a multilateral consensus—have played out on the global stage, notably in increasingly contentious elections over leadership of the International Telecommunication Union, a UN-affiliated specialized agency whose mandate encompasses internet regulation.[8] Similar fissures over domestic regulatory approaches toward AI are already evident and expected to bleed into nascent global governance initiatives. Managing normative and regulatory fragmentation in an eventual regime complex will thus be essential to advancing global AI cooperation.

Current proposals for the design of global AI governance have concentrated on several functions.[9] This paper consolidates these into four broad categories to facilitate deep analysis and probe the relevance of their associated institutional analogies. The first function is to provide an authoritative platform for scientific and technical knowledge and information sharing on the latest state of AI capabilities and their potential ramifications. The second is to promote common norms and standards for the responsible uses of AI by both public and private actors, as well as to seek to harmonize national regulatory approaches. The third is to support the broadest possible access to and equitable sharing of the benefits from AI, with a particular focus on the development needs of low- and middle-income countries. The fourth is to foster global collective security by creating frameworks to deter and respond to destabilizing uses of these technologies by state and nonstate actors, as well as to prepare for any existential risks posed by the potential emergence of artificial general intelligence.

Realizing these functional objectives will require a regime complex that is *multi-multilateral*, comprising several institutions and initiatives, each involving different membership groups. For some functions, building entirely new institutions may be necessary. More commonly, the mandates and capacities of existing institutions will need to be adapted and extended to make them AI-competent. Many institutions for AI governance will be intergovernmental, with membership restricted to sovereign states; some of these will have universal membership, whereas some will be narrower, selective, minilateral frameworks among like-minded nations. Other global arrangements will have multiple stakeholders, involving not only national governments but also corporations and civil society actors. Eventually, some normative commitments may become grounded in binding international law, while others will remain voluntary.

The Rise of AI and Efforts to Govern It

Breathtaking advances in AI and its integration throughout society have left public authorities scrambling to ensure the safety and transparency of its development and applications.[10] If the velocity of innovation accelerates, the governance challenges will become even more daunting.[11]

Although definitions vary, artificial intelligence generally refers to information-processing systems that use models and algorithms to make inferences from data and other inputs to "generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."[12] The main focus of attention today is on foundation AI models that are trained on large amounts of data to be utilized for a range of tasks, rather than on narrow applications.[13] These models have contributed to the increasing sophistication of generative AI, including an ability to produce high-quality text, images, and videos from input data and similarly respond to text and images, including through chatbots like ChatGPT.[14]

The potential benefits of AI—for alleviating poverty, transforming medicine, combating climate change, enhancing worker productivity, eradicating infectious diseases, improving access to high-quality education, strengthening the efficiency of local governments, and so much else—are significant.[15] Notably, however, these potential benefits have attracted less attention as subjects of international governance than AI's possible dangers. (This may reflect the underrepresentation in prominent initiatives of policymakers from countries of the Global South, where policy discussions have tended to focus more on the many opportunities that AI presents for development.)[16]

Instead, many policymakers and researchers in the Global North have emphasized AI's potential risks.[17] Among these concerns are that AI may facilitate political interference and the spread of misinformation and disinformation;[18] entrench discrimination through algorithmic

bias;[19] enable mass surveillance by authoritarian regimes;[20] worsen invasions of privacy by private corporations;[21] generate mass worker dislocation and unemployment in knowledge and data-intensive sectors;[22] exacerbate global inequality;[23] facilitate the spread of lethal autonomous weapons systems;[24] lower barriers to entry for biological and nuclear weapons;[25] weaken security in cyberspace;[26] accentuate geopolitical rivalry and the risk of major power war;[27] undermine nuclear deterrence and strategic stability;[28] and hasten the emergence of an omniscient, omnipresent, and omnipotent "superintelligence" that would act contrary to the interests of its human creators.[29]

Much of the action to regulate AI and manage its risks and opportunities will take place at the domestic level.[30] The EU, China, the United States, and India are already pursuing regulatory models, with different objectives and enforcement mechanisms. The EU's AI Act, provisionally approved in December 2023 and passed by the European Parliament in March 2024, seeks to protect "fundamental rights, democracy, the rule of law and environmental sustainability" without hampering innovation.[31] It adopts a "risk-based" approach, such that higher-risk AI applications face more stringent rules, especially regarding transparency and quality of data sources, cybersecurity, and safety testing.[32] This high-risk category includes foundation models for now, despite misgivings from France, Germany, and Italy about stifling innovation.[33] Banned applications include social scoring, emotion recognition in workplaces and educational institutions, and biometric categorization systems.[34] (Law enforcement can still use biometric identification systems for solving a narrow list of crimes.) The rules are legally binding, and noncompliant parties are subject to fines, though most provisions will not take hold until two years after the act enters into force.

China also imposes strict, binding regulations on companies—including with respect to specific components like algorithms, synthetically generated content, and generative AI.[35] Unlike the EU, its primary motivation is not to protect individual rights but to exert information control over Chinese society.[36] For example, China's Generative AI Measures, a set of regulations that took effect in August 2023, seek to restrict chatbots from "producing fake and harmful information," including content related to the "subversion of state power."[37] However, these efforts to contain politically sensitive content create a high regulatory burden for companies. Already Chinese policymakers have relaxed some rules and their enforcement, lest they stifle industry innovation and broader economic competitiveness.[38]

The United States is further behind in producing regulations. It has relied on voluntary commitments and nonbinding measures, undergirded by the principle of "responsible innovation."[39] In July 2023, President Joe Biden's administration reached agreement with seven leading AI companies on "voluntary safeguards," including security testing and monitoring of bias and privacy risks.[40] In October 2023, the White House issued the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, a notable first step in creating guardrails for advanced AI systems.[41] Still, the executive order primarily directs government agencies to prepare assessments and recommendations on AI

in their domains; robust U.S. regulation will require congressional legislation.[42] On balance, the current U.S. stance privileges innovation, consistent with a traditional deference to the private sector, whereas the EU prioritizes safety and rights.

Finally, India is seeking to carve out its own role in AI governance and promote an alternative regulatory framework for countries in the Global South that encourages domestic innovation while keeping citizens safe.[43] The form and content of the country's eventual AI regulations remain unclear, even though the Indian government's public policy think tank released a national AI strategy in 2018 and the Ministry of Electronics and Information Technology proposed the Digital India Act in 2023, which focuses on AI, data governance protection, and cybersecurity.[44]

Beyond these domestic regulatory steps, international governance will also be critical, given the global reach and ramifications of AI. Although the development of advanced AI systems and chips is currently concentrated in a handful of countries, access to many AI models cannot easily be contained within borders.[45] It is thus imperative to develop international rules of the road regarding AI's development and use, hold governments and private actors within their sovereign jurisdictions accountable for how they employ it, and create backstops in case governments are unable or unwilling to fulfill their regulatory responsibilities. The emerging global framework must also be consistent with established international law, norms, and conventions, including the UN Charter, the Universal Declaration of Human Rights, international humanitarian law, and other multilateral treaties.

Multilateral AI diplomacy has accelerated accordingly. In October 2023, at the Third Belt and Road Forum, Chinese President Xi Jinping announced a Global AI Governance Initiative.[46] Soon after, the G7 released the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems and a related code of conduct (hereafter called the Hiroshima Code of Conduct).[47] In November 2023, the United Kingdom (UK) hosted the world's inaugural AI Safety Summit, where twenty-eight nations agreed to commission an expert-led State of the Science Report, among other outcomes.[48] More recently, in December 2023, the HLAB released its interim report, "Governing AI for Humanity."[49] This flurry of activity builds on earlier work by intergovernmental bodies, such as the Organisation for Economic Co-operation and Development's (OECD) AI Principles; the UN Educational, Scientific and Cultural Organization's (UNESCO) Recommendation on the Ethics of AI, which was adopted by all 193 member states in 2021 (though the United States has not adopted the recommendation despite rejoining UNESCO in July 2023); and the Global Partnership on AI's multidisciplinary research reports.[50]

Notwithstanding these developments, any effort to govern AI at the global level will face powerful incentives working against such regulation, as both major powers and leading companies compete with their respective counterparts to reap the geopolitical and economic rewards of this new technology. Approaches to global governance for AI must therefore account for this barrier if they are to overcome it.

A Regime Complex for AI

The regulatory complexities presented by AI, as well as the multiplicity of actors involved and the geopolitical context, necessitate multiple institutions at the global level. Beyond the HLAB's interim report, which outlined a disaggregated global governance framework, other policy developments affirm this expectation.[51] For instance, the Bletchley Declaration, issued at the close of the UK AI Safety Summit, anticipates that countries will work together through "existing international fora and other relevant initiatives."[52] In a similar vein, the director of the White House Office of Science and Technology Policy noted in a lecture that "different [multilateral] fora are approaching different aspects of the [AI] problem."[53]

A regime complex is a collage of overlapping multilateral arrangements involving different actors, functions, and principles that facilitate international cooperation.[54] Similar arrangements have emerged to address other complicated global domains, such as climate change, global health, and cyberspace, though the cyberspace example serves more as a cautionary tale of how discord among major powers can produce extreme fragmentation in global governance. [55]

Key components of the regime complex for climate change, for example, include multilateral treaties, such as the UN Framework Convention on Climate Change and the Montreal Protocol on the ozone layer; scientific assessment bodies, notably the IPCC; funding mechanisms like the Green Climate Fund and the Global Environment Facility; UN agencies like the World Meteorological Organization (WMO) and the UN Environment Programme (UNEP); the climate initiatives of the International Monetary Fund (IMF), the World Bank, and regional multilateral development banks; narrower bodies like the International Energy Agency; minilateral groupings such as the G20 and the Major Economies Forum on Energy and Climate; networks of subnational actors like the C40 Cities coalition; and private sector coalitions such as the Glasgow Financial Alliance for Net Zero.[56]

Framing AI governance in terms of a regime complex is useful analytically because it draws attention to the varied practical imperatives and messy geopolitical realities of AI governance. This approach helps policymakers to break down the AI challenge into manageable chunks; anticipate the dilemmas of alternative institutional design choices, including the trade-offs between universal and coalitional approaches to regulation; and contemplate the relevance of analogous institutions that have managed other global challenges.

Regime complexes are nonhierarchical and modular, meaning that no institution holds authority over the other constitutive elements and that the constituent parts of a regime complex can be designed for specific purposes. Such a decentralized approach is well-suited to AI for three reasons.

First, AI governance must make simultaneous progress on several fronts, such as improving policymakers' understanding of AI's underlying science, ensuring wider access to the technology, and regulating its use in military contexts. Rather than asking a single institution to fill these needs, a division of labor is more appropriate, with different institutions pursuing cooperative objectives across distinct domains.

Second, using a variety of fora and institutions can permit selectivity in membership, allowing policymakers to adjust who is at the table, depending on the nature of the issue, the interests and competencies of relevant actors, and geopolitical considerations. Some aspects of AI cooperation warrant broad intergovernmental participation within institutions that feature universal membership like the UN and its agencies. Tackling other aspects may only be feasible within, or when restricted to, narrow coalitions of countries that share values and objectives, possess germane AI capabilities, or are able to move with greater dispatch than bodies with more encompassing membership. Still other frameworks will need to have multiple stakeholders, such that technology companies and civil society representatives hold formal membership alongside governments.

Third, regime complexes can advance governance in the absence of multilateral treaties or even formal organizations. Negotiating and ratifying international legal conventions is a painstaking process and will be impossible for many AI issues in the short and medium term. Instead, progress will initially rely on nonbinding agreements and declarations of principles and on the promotion of norms of behavior for states and nonstate actors, which can be gradually incorporated in the activities of existing and new international organizations. Regime complexes can therefore enable an adaptive, multifaceted approach to global governance that permits the iterative evolution of regulation in response to innovation.

Regime complexes have disadvantages, but highlighting them now can encourage policymakers to mitigate these downsides as they begin to build a governance architecture for AI.[57] First, the fragmented nature of regime complexes, particularly the absence of an authoritative institution or even high-level conductor to orchestrate actors and activities, can lead to incoherence, gaps, and redundancy across initiatives, undercutting progress on shared challenges and complicating efforts to hold governments accountable. To promote complementarity from the outset, countries must negotiate shared principles and norms about how to address the development and use of AI.

Second, regime complexes can exacerbate competitive dynamics among countries, providing nations dissatisfied with existing institutions greater leeway to engage in forum shopping or create alternative bodies that undermine the original ones. Such contested multilateralism, which is particularly common between strategic adversaries like the United States and China, can be corrosive to more encompassing forms of collective action, as nations prioritize narrow interests over shared goals.[58] While policymakers cannot eliminate competitive dynamics, they can temper them by supporting wider participation on some issues. In this sense, the UK's decision to include China at the November 2023 AI Safety Summit was prudent.

Form Follows Function

A regime complex is a multidimensional governance system that emerges not through the tightly coordinated actions of any single group of countries but through the cumulative efforts of disparate actors (including sovereign states, intergovernmental organizations, and nonstate actors) to construct institutions that address different aspects of a complicated global challenge. The future regime complex for AI will be no different. Despite this lack of central direction, policymakers in the United States and other major powers can help foster the emergence of an effective, stable, and coherent AI governance system if they focus on several concrete objectives and remain attuned to how these different initiatives and institutions can complement each other.

The regime complex for AI should notionally fulfill at least four main functions: building scientific understanding about AI's evolving capabilities and implications, setting standards for its development and use, sharing the benefits of AI globally, and promoting collective security. Given the multifaceted nature of these functions, the diversity of national interests involved, and the multiplicity of state and private actors in this field, the framework for AI governance that emerges in each functional area will likely rest not on a single institution but comprise myriad multilateral, minilateral, and multistakeholder arrangements.[59]

The following sections explore these four functions of a future regime complex for AI, as well as the relevance and limitations of prominent analogies to existing multilateral bodies in other fields.

Building Scientific Understanding

One institutional priority is establishing an authoritative intergovernmental framework for synthesizing and sharing the latest scientific and technological breakthroughs related to AI to give policymakers and the public a common baseline of understanding. Generating these reference points is a precondition for international cooperation on managing the risks and opportunities presented by advanced machine learning systems, most of which are developed by private corporations, with insufficient transparency and information sharing.

With this objective in mind, several recent proposals advocate the creation of an intergovernmental body to regularly assess, build political consensus on, and share the latest scientific and technical knowledge on AI's capabilities and implications, including its potential social, economic, developmental, environmental, political, and security-related impacts.[60] Such an expert-led process would ideally provide objective information to ensure that fact-based assessments undergird the development of national and international policies, including the formulation and harmonization of best practices.

At the UK AI Safety Summit, participating governments agreed to commission an international report on the state of AI science, under the direction of prominent computer scientist Yoshua Bengio, to better understand the power and risks of cutting-edge "frontier" models.[61] It is slated to be released before South Korea hosts a mini virtual AI Safety Summit in mid-2024. Although inspired by the IPCC, this initiative is currently ad hoc, lacking set procedures and bureaucratic infrastructure. Presuming this framework is formalized in a permanent institution, the IPCC and similar entities provide policymakers with some useful lessons. Still, significant differences between climate and AI challenges, as well as the shortcomings of the IPCC itself, could render this analogy less compelling than it initially appears.

Proposed Institutional Models

Established in 1988, the IPCC is an intergovernmental body under UN auspices, comprising 195 member states.[62] Its mandate is to produce scientific assessments of climate change, including current and future impacts. The IPCC elects a bureau of scientists to oversee each assessment report and select the experts who prepare and write that document. This complicated process involves input from thousands of scientists. (It is no surprise, then, that the IPCC's regular assessments only come out every six to seven years).[63]

Analogous bodies exist, including in other environmental fields. One is the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, created in 2012, which synthesizes the latest scientific findings on the status of and threats to biodiversity and Earth's natural ecosystems.[64] It has a structure similar to the IPCC, but it is not a UN body, and its mandate includes supporting policy development.[65] Another model is the Montreal Protocol (formed in 1987), whose three highly respected assessment panels address scientific, technical, and environmental questions related to the health and integrity of the ozone layer.[66]

Were one to envision a similar entity for AI built on existing initiatives, one practical question is under what auspices it should be created. The above examples suggest a range of options: the IPCC was established by the WMO and the UNEP. The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services was founded by a coalition of ninety-four governments at a conference in Panama in 2012.[67] The three ozone panels, meanwhile, were created pursuant to a formal multilateral treaty.

The IPCC offers an appealing governance model. First, it is ostensibly policy neutral, meaning it does not adopt a stance on actions that countries should pursue. Given the diversity of governments' regulatory approaches toward AI, a commitment to policy neutrality should minimize political contestation within a scientific assessment panel for AI and contribute to the panel's legitimacy, which is important due to low levels of trust toward major powers and private technology companies. Moreover, there is not yet sufficient research or agreement on AI's capabilities and impact for the proposed panel to advocate policy solutions.[68]

Second, the IPCC publishes periodic special reports (on subjects like the implications of global warming of more than 1.5 degrees Celsius or on the ramifications of climate change for Earth's oceans and frozen regions), which illuminate areas where future governance initiatives are needed.[69] For AI, there could be special reports on the extent of existential risk posed by AI or on the state of global investment in research and development on AI safety. Commentators from academia and industry have suggested creating an international AI research institution, along the model of the European Organization for Nuclear Research (CERN) or the International Space Station.[70] A similar endeavor for AI would be an expensive undertaking, so it should be preceded by an assessment of current primary research and whether there are gaps that an institution for joint research could fill.

Limitations of Proposed Models

As a research subject, AI is notably different from climate change, biodiversity, and the ozone layer, so establishing a scientific assessment panel for AI will involve addressing additional complexities and trade-offs. First, the rapid speed of AI innovation conflicts with the IPCC's painstaking, multiyear assessment cycles. IPCC assessments are designed to achieve a high degree of scientific rigor and solicit widespread participation at every phase.[71] Once authors write the chapters (based on primary literature in the field), people can register as expert reviewers to provide feedback on the initial drafts. The authors then write second drafts, submitting them for additional review by experts and governments. Final drafts are then prepared for governments, which submit their last edits and then meet to approve the report in question. This complex process has some advantages. By prioritizing rigorous research and inclusivity, it helps to build scientific consensus and political legitimacy for IPCC assessments.

However, a protracted timeline makes little sense for AI because the pace of innovation would render any report outdated by the time it appeared. AI requires a more agile approach to scientific assessment by continually evaluating the technology's evolving capabilities and their ramifications. Streamlining the assessment process for AI will also require policymakers to make difficult trade-offs between inclusivity and efficiency, particularly when it comes to who determines who should participate in these assessments and how. To reflect the entire range of AI risks and opportunities, any intergovernmental assessment would ideally draw on experts from countries at all income levels and from a wide range of disciplines, including non-technological fields such as ethics, law, and the social sciences. At the same time, such a body will need to move with dispatch. One plausible scenario would be to create separate panels to assess specific dimensions of the AI challenge, analogous to the three panels created under the Montreal Protocol, for a more efficient division of labor. In addition, policymakers should consider including a horizon-scanning function that alerts the international community in real time to emerging dangers and dilemmas, akin to what the Financial Stability Board does for the global financial system.[72] This function should include regularly shared risk assessments based on an agreed-upon risk classification system.

Second, policymakers need to decide on a governance model for the assessment body or bodies for AI and determine the precise role of the private sector. Many representatives from governments, technology companies, academia, and civil society endorse a multistakeholder approach to AI governance, without specifying how this objective should influence the institutional design of any future assessment panel. The IPCC is formally intergovernmental; it is a UN body, and member states oversee how it functions and approve the final report. At the same time, it has multistakeholder dimensions: its authors and reviewers participate in their individual capacities, are nominated and selected based on their expertise, and can come from industry and civil society, not simply academic or government research institutions.[73] For an AI assessment panel, is this the model that those who advocate a multistakeholder approach have in mind? Or should the oversight and governance of an AI panel also be multistakeholder, and, if so, how would that be designed?

The answers could have important implications for AI regulation. Since private industry leads the research and development of advanced AI, any scientific body needs to negotiate standing arrangements with major technology platforms to gain access to and information about the latest models. In addition, private sector actors presumably need to be represented (and perhaps heavily so) in the authorship of assessment reports. But if the panel's purpose is to provide policymakers with information that will, in part, contribute to regulatory decisions, then governments may want to maintain ultimate oversight.

Finally, a core objective of any scientific assessment body should be to advance transparency in AI's research and development. Accordingly, governments should assign to this body a role as a registry or clearinghouse for up-to-date information about advanced civilian AI research and development, including any recent breakthroughs in capabilities. (Such information sharing is inherently more challenging when it comes to potential miliary applications, as discussed later.) This clearinghouse role would help compensate for any time lag between the body's periodic assessments, allow the world to keep abreast of rapid AI innovations, and facilitate multilateral cooperation by reducing uncertainty. Analogous proposals have been made in other fields involving transformative or unconventional technologies, such as solar geoengineering. Some of them have even been implemented, notably the Biosafety Clearing-House created under the Cartagena Protocol on Biosafety to share information on genetically modified organisms.[74]

At present, no single international institution exists that can be adapted to fulfill the governance role of building scientific understanding on AI. The twenty-nine-nation Global Partnership for AI, officially launched in June 2020 with sponsorship by Canada and France and now chaired by India, was initially envisioned as an IPCC-like body but has not yet achieved an authoritative role.[75] It is unlikely to do so in the near term, in part because membership to the partnership is contingent on endorsing the OECD AI Principles, limiting its expansion to more countries. A more likely path to a scientific body—reflected in the HLAB interim report—may be to create a new standing institution out of the ad hoc effort to produce the report on the state of AI science.

Setting Standards and Harmonizing Regulations

A second core function of a regime complex for AI should be to establish common standards for its responsible development and use by state and nonstate actors and to harmonize the regulations emerging from domestic jurisdictions. This should eventually include mechanisms to monitor the implementation of and verify compliance with agreed-upon standards, even if they are voluntary and nonbinding.

Recent declarations to promote international AI principles—emerging from political platforms as diverse as the G7, OECD, UNESCO, G20, China's Ministry of Foreign Affairs, and the UK AI Safety Summit—employ the same assortment of high-minded adjectives.[76] "Ethical," "responsible," "trustworthy," "human-centered," "transparent," "safe," "accountable," and "fair" are prominent examples. This consensus suggests that these high-level principles are relatively settled, until they need updating to reflect advancements in AI. Yet this unanimity on the surface about aligning AI with human values belies the fragmentation already characterizing domestic regulatory approaches. It also overlooks the divergences in values among some major powers, which will likely increase the attractiveness of parallel minilateral forums, in which narrower like-minded coalitions can push for higher-quality or alternative standards.

As more countries develop and implement domestic rules and frameworks to manage AI risks and opportunities—beyond major powers like the EU, China, the United States, and India—pressure will grow to translate existing high-level multilateral principles into common standards and harmonize disparate national regulations, as the HLAB interim report highlighted. One jurisdiction's regulatory approach is unlikely to become a de facto global benchmark, given the pluralism of current preferences.[77] In the meantime, some experts have recommended creating a new UN-affiliated specialized agency to address this burgeoning regulatory fragmentation.[78]

Proposed Institutional Models

Two specialized UN agencies often invoked as models for harmonizing AI standards and benchmarks are ICAO and the International Maritime Organization (IMO). Each provides an intergovernmental forum to set regulatory and technical standards for a specific domain and encourage their implementation. Established in 1947 after the entry into force of the 1944 Convention on International Civil Aviation, ICAO promotes safe international air transport, including by setting technical standards and developing best practices, monitoring their domestic implementation, and supporting aviation capacity building, based on assessments prepared by representatives and industry experts from the body's 193 member states.[79] IMO, which opened its doors in 1958 after its 1948 convention went into effect,

fosters cooperation among its 175 member states on international shipping, with a focus on developing standards and providing technical guidance on the implementation of measures related to issues like safety, security, pollution, and efficiency.[80]

The idea of designing a similar multilateral agency to set global AI standards and harmonize domestic AI regulations has proven attractive. First, like civil aviation and shipping, AI is a transnational phenomenon with cross-border effects. To ensure AI safety, security, and efficiency, both the private sector and national governments need to conform to a set of minimum global standards in AI's development and use. Like airlines and shipping companies, AI technology companies operate across domestic jurisdictions, and they rely on interoperable standards for their own self-governance, policy development, and supply chains. Regulatory harmonization on AI is also imperative for governments. It would help ensure that high-standard jurisdictions are not placed at a competitive disadvantage compared to low-standard ones. Such harmonization would reduce opportunities for companies to engage in regulatory arbitrage by gravitating to jurisdictions with lax standards, and it would discourage the emergence of regulatory black holes that could be exploited by nefarious state and nonstate actors.

Second, the ICAO/IMO model also includes supervisory and monitoring mechanisms to track compliance and ensure accountability. ICAO, besides providing robust guidance on the implementation of standards, conducts safety and security audits on the actions, including legislation, that countries are taking to implement standards.[81] (It does not audit industry actors.) The results of safety audits are made public, whereas security audit results remain confidential among member states. In cases of noncompliance, states can use dispute settlement mechanisms to suspend the voting power of another state and, in principle, impose UN General Assembly or UN Security Council sanctions.[82] Additionally, if an airline is found to be violating rules, then member states are all expected to prohibit it from operating in their airspace. By contrast, IMO has a softer monitoring role. Although audits of national implementation of standards became obligatory in 2016 (after previously being voluntary), IMO lacks a dispute settlement mechanism or even minimal enforcement provisions.[83]

Limitations of Proposed Models

At the same time, the ICAO/IMO model has significant limitations. First, AI is not limited to a single domain but is rather a general-purpose technology with the potential to affect all aspects of society, akin to electricity. The world thus needs multiple sets of standards to address both AI's technical dimensions and its development and use in various domains. Rather than seeking to create a single, new, all-encompassing standard-setting body to govern AI's application to an ever-expanding set of specific use cases, the goal must be to make existing international institutions as AI literate as possible, as soon as possible. This will require promoting international dialogue on domain-specific AI standards so they can be developed and implemented in particular fields.

Much global standard setting is highly technical, focused on the establishment of universal benchmarks, guidelines, measures, or models that facilitate international coordination.[84] Classic examples include the establishment of standard time, mutual recognition of internet protocol (IP) addresses generated by the Internet Assigned Numbers Authority, or the creation of standards for capital adequacy determined by the Basel Committee on Banking Supervision.[85] As AI permeates the global economy, demand will increase for common standards in its applications, as well as guidelines for assessing its risks and testing the effectiveness and quality of AI systems.

This technical work will be relatively straightforward, focused on the interoperability of technologies across jurisdictions.[86] Some of these details are already being worked out under the auspices of the International Organization for Standardization, a nongovernmental entity comprising 170 national standards bodies that seeks to forge international consensus on market-relevant global standards.[87] Other "sector-specific," or "vertical," standards are being produced through the more narrowly focused International Electrotechnical Commission, though the commission is also collaborating with the International Organization for Standardization to create "generic," or "horizontal," standards.[88] Government entities, including the U.S. Commerce Department's National Institute of Standards and Technology, are also ramping up these efforts.[89]

In parallel with technical standard setting, the world needs sector-specific international standards for the development and use of AI applications, which can inform governments as they develop national regulatory policies. This is analogous to what is already happening at the national level. For instance, the White House's October 2023 Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence directs individual U.S. government agencies and departments to come up with guidelines on the application of AI in sectors as diverse as employment, financial services, healthcare, and transportation.[90] At the multilateral level, too, existing institutions are beginning to develop sector-specific guidelines. The IAEA is starting to work on the application of AI in nuclear science, power, safety and security, and verification, while the World Health Organization (WHO) has begun to develop standards for use of AI in global health. ICAO could likewise be tasked with catalyzing standards for use of AI in civil aviation, building on EU efforts.[91] To gain diplomatic traction and exert practical effects, such standard-setting efforts should not be limited to the purview of agency secretariats; they should also involve dialogue among member states.

Second, the ICAO/IMO model is intergovernmental rather than multistakeholder. This is potentially problematic in the case of AI because the private sector dominates the design, development, and distribution of the technology and its applications. Although the phrase global governance conjures images of formal intergovernmental organizations, in practice many global regulatory and standard-setting bodies are multistakeholder.[92] Constituting an organization in this way would grant representatives of the private sector and civil society organizations a role alongside participants from governments and international agencies in the processes of agenda setting, negotiating, implementing, monitoring, and enforcing or evaluating relevant standards and regulations. This approach may hold lessons for AI.

Among the most prominent examples is the International Labour Organization (ILO), established in 1919 and now with 187 member states.[93] It sets labor standards through conventions and nonbinding agreements and provides technical assistance, helping countries draft domestic legislation to meet international obligations.[94] Unlike ICAO and IMO, the ILO has a tripartite governance structure involving governments, workers, and employers.[95] Each member state delegation to the ILO's governing body, the International Labour Conference, must include an employer and worker representative, alongside two representatives from the country's government.[96] Another prominent multistakeholder regulatory and standard-setting entity is the Internet Corporation for Assigned Names and Numbers, an independent nonprofit charged with coordinating maintenance of the internet's Domain Name System of unique identifiers.[97] The nonprofit's Governmental Advisory Committee comprises representatives from 179 sovereign nations and dozens of international organizations. Still, as these examples show, adopting a multistakeholder design is not a complete solution. For AI, a key challenge would be determining the degree and form of representation for each type of party (including governments, the private sector, and civil society).

The third major limitation of the ICAO/IMO model is that it is grounded in a near-universal membership framework, by way of international treaty law. As such, it is not well-suited to current geopolitical conditions, particularly the resurgence of strategic rivalry and ideological competition between the democratic West and authoritarian China and Russia, to say nothing of the hurdles to treaty ratification (not least in the United States). In this context, high-level intergovernmental standard-setting efforts for AI are likely to unfold along at least two parallel tracks: all-encompassing UN-affiliated settings and narrower coalitions of like-minded participants.

The UN and its agencies will inevitably foster global dialogue on evolving principles, norms, and standards for the responsible use of AI and will promote greater transparency and harmonization of national regulations. Historically, the global body has played an important norm- and standard-setting function on topics ranging from sustainable development to the prevention of atrocities, drawing on the unparalleled legitimacy conferred by its universal membership and the binding UN Charter.[98] The current UN secretary-general, Guterres, has tried to do the same for AI. The HLAB he appointed is slated to prepare its final recommendations for consideration by the UN General Assembly at the Summit of the Future in September 2024. Although "the UN cannot and should not seek to be the sole arbiter of AI governance," as the HLAB acknowledges, the General Assembly can help promote coherence by negotiating and passing a declaration of principles for the development and use of AI.[99] Properly crafted, such a resolution could play a role similar to the nonbinding Universal Declaration of Human Rights, which laid the foundation for subsequent international legal conventions.[100] Furthermore, the General Assembly can ensure that standards are consistent with, and where possible take direction from, established international law and norms. Still, current geopolitical fragmentation suggests that any universal approach is likely to reflect lowest-common-denominator outcomes and hardly go beyond UNESCO's Recommendation on the Ethics of AI, adopted by all 193 member states in 2021.[101]

Beyond endorsing principles and norms, the UN can foster regulatory harmonization through its technical agencies. One promising approach would be to require that countries self-report domestic AI regulations to an international clearinghouse, much as trading nations are obligated to declare subsidies to the World Trade Organization.[102] A possible repository of this information is the International Telecommunication Union, the world's oldest intergovernmental organization, which already coordinates and reports on AI-related activities conducted across UN entities.[103] This notification process could support transparency and eventually lead to greater multilateral agreement on rules.

While it is worth pursuing universal approaches, the UN's standard-setting and regulatory role is still expected to be limited. Ideological differences among major powers may thwart meaningful global consensus, perhaps making it so that resolutions on AI norms and standards reflect the lowest common denominator. As countries translate their stated commitments into actionable standards and regulatory schemes, fissures between open and closed societies are likely to loom large and at times be insurmountable. For example, the advanced G7 market democracies are unlikely to concur with authoritarian China and Russia on the "ethical" or "human-centered" standards that should inform the use and export of AI tools for mass surveillance and censorship, or even on how to reduce algorithmic bias. (China and Russia signed on to UNESCO's Recommendation on AI Ethics, which states, "AI systems should not be used for social scoring or mass surveillance purposes," but this is only a voluntary commitment, lacking an enforcement mechanism.)[104]

Accordingly, the community of advanced market democracies may well pursue parallel minilateral efforts to adopt higher regulatory standards and more demanding monitoring provisions. For instance, the G7 could set up a body to assess national implementation of the nonbinding Hiroshima Code of Conduct among endorsing countries. In this scenario, the G7 would treat any emerging UN standards as a floor while separately pursuing higher standards—with the aspiration to eventually globalize the latter. The United States and other Western nations are no strangers to this approach, having created high-ambition coalitions in other domains and then subsequently inviting other countries to join. The Proliferation Security Initiative, the Artemis Accords, and the Declaration for the Future of the Internet are notable examples.[105]

Among the most successful such ventures is the Financial Action Task Force (FATF).[106] In 1989, the G7 established the FATF to combat money laundering, and the body's mandate was later expanded to combat the financing of terrorism and of the proliferation of weapons of mass destruction. Over time, the FATF has developed widely accepted standards that it uses to monitor the strength of countries' legal and regulatory structures and classify countries as cooperating or noncooperating jurisdictions, depending on whether they implement sound practices. Although this is an informal arrangement, the FATF's designations have real bite, as private financial institutions tend to reduce their exposure to blacklisted

countries. The IMF and the UN Security Council have since accepted FATF standards, demonstrating how a club of like-minded nations can elevate overall global norms and standards. An analogous framework for AI could similarly classify and rate jurisdictions to discourage private corporations from operating or investing in low-standard jurisdictions, limiting their economic prospects.[107]

One risk in the minilateral approach is that rival coalitions—for instance, the BRICS (Brazil, Russia, India, China, and South Africa) or the Shanghai Cooperation Organization—could mimic this method, accelerating the fragmentation of the global economy and world order. A related danger is that such an approach could deepen divisions between the Global North and the Global South, if it came across as another attempt by privileged powers to impose their standards on poorer players. To avoid this hazard, advanced market democracies need to collaborate with democracies in the Global South in the development of AI standards and regulatory harmonization. India, which has staked out a distinctive regulatory approach for AI that balances opportunity with rights, will be an important partner in any such effort.[108]

Finally, Western countries will have to overcome some of their own divisions if they are to advance common standards, an obstacle exemplified by ongoing debates about the Council of Europe's Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law.[109] The draft treaty—which if finalized would be the first binding international AI convention—offers the West an opportunity to present a united front. Yet that outcome depends on whether negotiators can resolve big disagreements among both the council's forty-six member states and observer nations like the United States, including over whether the treaty applies to the private sector.[110]

Ultimately, an effective regime complex for AI needs to include mechanisms for monitoring and verifying compliance with standards negotiated in multilateral or minilateral settings. While a robust global regulatory scheme grounded in international law remains a distant prospect, lessons from other domains suggest that well-designed frameworks can help governments verify and improve compliance with nonbinding standards. Examples include the UN Human Rights Council's process of Universal Periodic Review, through which countries assess each other's human rights records; the obligation of ILO members to submit periodic self-assessments of compliance to the International Labour Conference; and the Enhanced Transparency Framework under the Paris Agreement, which mandates that countries, starting in 2024, report on the mitigation and adaptation measures they have taken to address climate change.[111] Similar self-reporting and peer review arrangements for AI standards may be feasible even for countries that are not inherently like-minded.

Sharing Access and Benefits

A third function of the regime complex for AI should be to expand access to and the sharing of benefits from this technology, as its development and use remain concentrated in a few advanced economies. About 2.6 billion people, or approximately one-third of the global population, are still unconnected to the internet, so ensuring that all countries and citizens can benefit from AI will be no small feat.[112] Yet doing so is critical. The world is badly off track in terms of meeting the UN Sustainable Development Goals (SDGs), a set of globally agreed-upon objectives for bettering the human condition by 2030, and AI has the potential to alleviate or exacerbate this reality.[113]

The applications of AI for achieving the SDGs appear almost limitless, including AI's potential to deliver high-quality education and healthcare services, advance early warning systems for extreme weather events, make agriculture more climate resilient, and support biodiversity and ecological conservation.[114] Yet this outcome is not guaranteed. The countries leading in AI development have a moral obligation, and practical incentives, to share its benefits. Supporting critical investments in the Global South is a start, since across these otherwise heterogenous countries, AI has been perceived as a beacon of opportunity. (This view contrasts with the focus on risk in the Global North.)[115]

Traditional development actors—including intergovernmental organizations, national development agencies, and nongovernmental organizations—are already using AI applications to improve service delivery, especially for SDG-related projects. For example, the UN Development Programme has leveraged AI tools to identify trends in hate speech in Sudan and electoral misinformation in Zambia and Honduras, strengthen government policy evaluation in Mexico, and accelerate cash transfer processes for beneficiaries in Togo and Bangladesh.[116] The UN Office of the High Commissioner for Refugees has used AI-based predictive analytics to improve the agency's responses by forecasting movements of internally displaced people in Somalia and of Venezuelan refugees and migrants into Brazil.[117] The U.S. Agency for International Development has also funded projects for machine learning applications in sectors like healthcare, democracy and governance, humanitarian response, and education.[118]

More ambitious capacity-building efforts are needed to promote broader access to AI technology itself, not just the delivery of benefits derived from its use. As the HLAB's preliminary report highlighted, this entails distributing to low- and middle-income countries (LMICs) key AI inputs like computational power (compute)—the hardware and software needed to build new models—data, and existing models.[119] Supporting training for public and private sector workers on the development and use of AI applications is also essential. One recent initiative in this mold is the UK-led AI for Development Programme with the United States, Canada, and the Bill and Melinda Gates Foundation. This coalition has committed to funding $100 million in programs across the African continent to "support home-grown AI expertise and computing power" and "solve some of the developing world's

most pressing challenges."[120] An ancillary goal is to support the creation of "sound regulatory frameworks for responsible, equitable, and safe AI." Such a comprehensive approach is critical to combating the digital divide and ensuring that developing countries and their citizens gain AI capabilities.[121]

Among other objectives, this agenda for cooperation on AI development should seek to promote joint science and technology research, invest in capacity-building initiatives for entrepreneurs and programmers creating new models and applications, and promote skills training for workers whose economic prospects are likely to depend on AI literacy. Such cooperation should also enhance data collection programs so that models are trained on data that are more culturally, linguistically, and geographically diverse, and it should support governments as they design domestic regulations and increase their own use of AI applications.[122]

To be sure, AI capacity-building strategies need to be tailored to specific national contexts, since any society's ability to leverage AI depends, in part, on situational factors like government effectiveness, digital infrastructure conditions, and human capital levels.[123] Still, the nations that lead in AI capabilities ought to make an ambitious, formal commitment to share the benefits of AI, including as part of the Global Digital Compact that UN member states are currently negotiating.[124] This action is particularly prudent and warranted, given the impression in many developing countries—one compounded by the experiences of the COVID-19 pandemic and the deepening global climate emergency—that wealthy nations are fickle partners in tackling shared challenges and indifferent to the plight of the world's less fortunate.

Proposed Institutional Models

Two broad institutional models have been proposed to help expand access to AI technologies and financing for them.[125] One is modeled on global health partnerships and another on existing arrangements for peaceful uses of nuclear energy. Among other differences, they diverge on the degree of conditionality, if any, that should govern access to relevant financing, technology, and applications. The ultimate institutional framework for AI-related benefit sharing will likely fall in between these two templates, with moderately conditional access to AI.

The first model is based on public-private partnerships in global health, in particular the Global Alliance for Vaccines and Immunizations (Gavi) and the Global Fund to Fight AIDS, Tuberculosis, and Malaria (hereafter the Global Fund). This model suggests a potential way to unlock financing and make it possible for LMICs to access markets and innovative technologies.

Scarce financing presents a major hurdle to developing nations' ability to obtain equitable benefits from breakthroughs in AI. Beyond placing existing technology out of reach for many LMICs (including when it comes to licensing), inadequate financing constrains the

ability of private corporations in developing countries to obtain inputs for creating new applications and potentially their own models. This scarcity of financing also limits the capacity of LMIC governments to expand digital public infrastructure and launch upskilling programs so that their citizens can take advantage of the AI revolution. Complicating matters further, the profit incentives of leading AI companies may not steer them toward designing applications that address problems predominantly affecting LMICs. For these reasons, some researchers have cited public-private partnerships for public health interventions in developing countries as possible models for expanding global access to AI.[126]

Gavi was founded in 2000 by the Bill and Melinda Gates Foundation, the WHO, the UN Children's Fund (UNICEF), and the World Bank to expand immunization programs in low-income countries, particularly for diseases like malaria, pneumonia, and rotavirus.[127] Gavi continues to make vaccines more affordable for countries by negotiating prices with manufacturers or sharing the costs with governments. For the malaria vaccine, Gavi's advanced funding commitments ensured that manufacturers continued production.[128] As a funding mechanism, Gavi's country-based operations are conducted by domestic health ministries, alongside the WHO. Founded in 2002, the Global Fund is a similar form of collaboration among international organizations, philanthropic foundations, and national governments.[129] It allocates funding based on proposals and implementation plans submitted by countries after multistakeholder consultations.[130] The fund itself does not maintain an in-country presence, instead relying on existing national and international public health actors.

A comparable funding and partnership approach for AI offers a potential pathway for developing countries—and domestic private sector actors—to obtain access to existing AI products (many of which are developed with proprietary technology). Such an approach could also support the development of indigenous AI models, applications, computing capabilities, and human capital. Moreover, this type of financing mechanism would afford partner countries autonomy in the design and implementation of country-specific programs, allowing host governments to deploy such capacity-building resources as they see fit. (Some might prioritize developing digital public infrastructure or enhancing data collection, whereas others might prioritize accessing computing power for domestic technology firms.) Finally, this multistakeholder model would involve a role for international development agencies, multilateral development banks, governments of donor countries and developing countries, the private sector, and both global and domestic civil society actors.

However, unlike the provision of vaccines and medicines, the introduction of AI technologies and systems could have some destabilizing consequences. International donors are thus likely to insist that initiatives to spread this technology be accompanied by safeguards. An alternative model is a conditional access approach that includes provisions to guard against misuse, loosely analogous to the peaceful uses of nuclear energy pillar of the Treaty on the Non-Proliferation of Nuclear Weapons (NPT).[131] Like the public health model, it recognizes the importance of expanding access and benefit sharing but with the caveats that beneficiaries must show restraint and conform to other obligations (such as abstaining from the

pursuit of nuclear weapons). For AI, a contingent access framework would require beneficiaries of relevant technologies, applications, compute, financing, and other tools to conform to safety standards and refrain from certain kinds of development and use.

The NPT is based on a core bargain, whereby non-nuclear-weapons states agree not to acquire such weapons in return for a pledge by the five acknowledged nuclear-weapons states to pursue nuclear disarmament and share the benefits of access to peaceful nuclear technology.[132] As a model for sharing AI's benefits, the NPT's most relevant element is Article IV, which establishes an "inalienable right" of all parties to "develop research, production and use of nuclear energy for peaceful purposes."[133] That article further commits treaty parties, especially nuclear-weapons states, to ensuring that all nations have access to the equipment, materials, and scientific and technological information required to pursue "nuclear energy for peaceful purposes . . . with due consideration for the needs of the developing areas of the world." For more than five decades, Article IV has provided an international legal foundation for the transfer of nuclear technology and material to NPT member states to develop safe civilian nuclear energy programs, contingent on having safeguards that meet IAEA standards—to ensure that these inputs are not being diverted to nuclear weapons programs.

In recent years, the IAEA has enhanced its financing and technical assistance for this conditional access regime, launching the Peaceful Uses Initiative in 2010 to secure extrabudgetary funding for peaceful use projects and the COMPASS program in 2020 to support capacity building on safeguards implementation.[134] In addition, the IAEA's nuclear fuel bank provides low-enriched uranium to member states in circumstances where they are unable to secure it, provided safeguards are in place.[135] The IAEA also maintains a Global Nuclear Safety and Security Network, a knowledge- and capacity-building hub for countries with limited nuclear energy programs, so that those programs remain secure and in line with standards.[136]

Limitations of Proposed Models

When it comes to benefit sharing, the regime complex for AI is expected to fall somewhere along the continuum of unrestricted to restricted access. Countries leading in AI innovation will seek to balance the development imperative with safety concerns, tying financing for capacity building with regulatory support, as is the case for the UK's AI for Development Programme. This will likely entail that recipient countries agree to certain safeguards, though what these should look like, how they should be established, and how stringent they should be all will be matters for debate. At the very least, safeguards will not be as strict or robust as those associated with the NPT because nuclear energy has a clearer link to a catastrophic weapon than AI does with any potentially destabilizing uses.[137] Beyond this general point, AI presents challenges very different from both nuclear energy and global health. These models cannot simply be replicated as policymakers seek to design an AI benefit-sharing framework.

First, the success of the NPT model owed much to the dynamics of mutually assured destruction, which persuaded the world's two main nuclear powers—the United States and the Soviet Union—to pursue nonproliferation and arms control, including by curbing the availability of nuclear material. With no alternative provider, countries seeking access to peaceful nuclear energy were forced to accept stringent conditions for access.

It is uncertain today whether China and the United States, given their intense geopolitical and geoeconomic competition, could reach agreement on establishing parameters for granting developing countries access to advanced AI models and applications, or even reach consensus on what aspects of these technologies could be destabilizing. And if they cannot, the United States and other Western countries will need to think carefully about whether to impose their own strict conditions, given the risk that this could push countries toward China as an alternative supplier. (Control of AI's broad suite of technologies—and the knowledge associated with their development and use—is also less straightforward than control of nuclear weapons, a challenge further elaborated on in the following section on collective security frameworks.)

More generally, disagreements between major powers in the standard-setting sphere could spill over into the design of benefit-sharing arrangements, resulting in a fragmentation of approaches to governing access. Divisions between democratic and authoritarian powers are likely to be especially salient, given the different approaches of Western countries and China toward development cooperation and aid conditionality. Whereas Western donors continue to condition their assistance on commitments to good governance, human rights, sustainable environmental policies, and the like, China has sought to distinguish itself through its no-strings-attached (albeit mercantilist) stance, not least through its spree of financing for infrastructure projects throughout the world by way of the Belt and Road Initiative.

With AI, China will presumably commit to respecting countries' sovereignty over how they employ these technologies. Indeed, Chinese surveillance technology, empowered by AI, is spreading abroad.[138] Western nations, by contrast, are more likely to insist that partner nations develop and use AI in ways that support democracy and human rights. Once again, however, they must walk a fine line, seeking to incentivize rights-based AI governance while recognizing that strict conditionality may drive other countries into China's embrace.[139]

A second limitation is that the eventual global framework to finance and support AI access and benefit sharing must be designed to be inclusive, equitable, and non-extractive. Current AI models and related algorithms are not always trained on globally representative data, which can limit their utility and appropriateness in contexts beyond where they are developed.[140] This is an issue that broad access approaches do not address.

Capacity-building efforts to support the development of AI models and applications in more countries will help compensate for this problem, but the emergence of such indigenous capacity will take time. Major technology companies in wealthy nations will continue to drive much AI development, and these firms will need to gain access to more diverse data

to develop better and more accurate models and applications. This process is inherently extractive, even if the end result winds up better serving the interests of populations in developing nations. This dynamic poses an ethical dilemma—how to compensate the people of developing nations for the use of their data, so that they can share in the economic and other benefits derived from its use (benefits that may otherwise be concentrated in the nations where major technology companies are based).

One institutional precedent worth exploring is the Nagoya Protocol on Access and Benefit-sharing, which was signed in 2010 and entered into force in 2014 as a supplementary international agreement to the Convention on Biological Diversity.[141] The protocol was designed to ensure the fair and equitable sharing of benefits arising from the use of genetic resources, as well as traditional knowledge associated with them. The protocol establishes state obligations with respect to access to and use of these resources (including rules for prior informed consent and the issuance of permits), benefit sharing (including compensation in the form of royalties, knowledge, or technology transfer), and compliance.[142] Like many multilateral treaties, it seeks to balance the interests of developing and developed nations, in this case by ensuring that countries and communities in the Global South, where much biodiversity is located, share in the material benefits from any commercial and other exploitation of genetic resources, a domain dominated by companies from the wealthy Global North.

The Nagoya Protocol provides one possible model for a future framework for AI data governance that would allow national governments, particularly of developing countries, to regulate data harvesting by foreign technology companies and related organizations. Such an effort would admittedly be complicated, in part because while natural resources are considered a sovereign resource under the Convention on Biological Diversity, there is less agreement that data possess the same status.

A third dimension left out of broad access models is the role of joint scientific research. The countries currently leading the development of AI can help improve access and benefit sharing by launching a robust program of cooperation between the Global North and the Global South on scientific research, designed to support capacity building in countries that are hungry to translate technological innovation into development outcomes. This cooperation could seek to replicate existing international models of joint scientific research, such as CERN, the world's largest particle physics laboratory, or the International Thermonuclear Experimental Reactor, which seeks to build the world's largest fusion device. In this case, however, the emphasis would be on collaboration between developed and developing countries, similar to the World Climate Research Programme under the auspices of the WMO.[143] This approach could provide another way to incentivize agreement on safe, rights-based standards and promote access, without imposing strict conditions.

A final point merits emphasizing. Ensuring AI access among countries of the Global South is not the same as ensuring participation by those same countries in the global governance of AI, and promoting one of these objectives does not guarantee the other. One opportunity to enhance the influence of developing nations in global AI governance is in the Global

Digital Compact, a set of "shared principles for an open, free and secure digital future for all," slated to be submitted to the UN General Assembly for approval at the 2024 Summit of the Future.[144]

Promoting Collective Security

The fourth function of any international regime complex for AI should be to promote collective security as the proliferation of AI reshapes this domain, amplifying existing risks and introducing new, potentially catastrophic, ones.[145] Given the fractious geopolitical environment and competing interpretations of what constitutes collective security—and how AI will affect it—pursuing this mandate will require diverse mechanisms that seek to deter the malicious use of AI, including by mitigating dual-use risks; prevent a destabilizing AI arms race, through confidence-building measures (CBMs) and other means; and create safeguards, trip wires, and contingency-planning mechanisms to address emerging threats.

To start, breakthroughs in AI have turbocharged long-standing debates about lethal autonomous weapons systems (LAWS), as these systems transform military affairs.[146] Having reached the battlefields in the wars in Ukraine and Gaza, LAWS are increasingly contributing to use of force decisions and actions, and other AI applications are becoming integrated into broader functions like command and control; surveillance, intelligence, and reconnaissance; logistics and training; and information management.[147] The arrival of fully autonomous systems portends a paradigm shift in warfare, as countries employ systems that can "identify, track, and prosecute targets without human oversight."[148]

Despite fervorous multilateral diplomacy to negotiate legally binding restrictions on LAWS, a treaty that bans or severely limits the use of autonomous weapons is not realistic at present.[149] The United States and other major military powers have resisted efforts to restrict the use of LAWS, instead forging ahead with developing these systems, which they deem critical to lessening the destruction of war and, implicitly, maintaining their own military competitiveness. Given these trends, efforts to reach agreement on basic principles of use, especially regarding how existing international humanitarian law applies, hold more promise than a formal arms control treaty.[150]

While the development and deployment of LAWS and other military applications of AI are reshaping the battlefield, the challenges that AI poses to global collective security are far broader.

An effective regime complex for AI will need to encompass higher-level institutional frameworks that allow the international community to deter and respond to potentially destabilizing uses of these technologies by state and nonstate actors—and perhaps by AI itself.

AI threatens to undermine collective security in at least three ways, beyond LAWS. First, AI will increase the availability and lethality of all weapons—including weapons of mass destruction.[151] It will make it easier for state and nonstate actors to design novel pathogens for biological warfare, turn drugs and compounds into more effective chemical weapons, and build more powerful and precise nuclear weapons. The AI revolution will also enable more sophisticated malware, leading to more potent cyber attacks on civilian and defense infrastructure.

Second, the pace of AI innovation will exacerbate geopolitical competition, as major powers engage in an arms race to dominate this technology and translate these advantages into military supremacy.[152] Automated warfare, combined with AI-enabled disinformation, could increase the risk of major power conflict due to escalation, miscalculation, loss of command and control, or algorithm-determined retaliation to a preceding attack.

Third, the rapid advancement of AI capabilities could pose serious, even existential, risks to our species. Such a scenario may seem far-fetched and remains hypothetical, but intense technological competition to develop cutting-edge models and applications could generate selection pressures for selfish AI traits, analogous to biological evolution.[153] In principle, these dynamics could encourage the emergence of super intelligent AI that acquires powerful capabilities and uses manipulation to pursue objectives misaligned with its creators' intentions or survival.

Unfortunately, incentive structures may work against robust international cooperation to address these threats, as both governments and corporations compete with counterparts to reap the technology's rewards. Zero-sum thinking may lead major AI powers to prioritize short-term gains over long-term global peace and stability, exacerbating the security dilemma inherent in world politics.[154] Cutthroat commercial competition could likewise encourage leading AI companies to sacrifice safety for innovation—a risk Silicon Valley itself recognizes. In March 2023, 1,000 technologists and other experts advocated a six-month pause in "the training of AI systems more powerful than GPT-4," warning that private labs were "locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one—not even their creators—can understand, predict, or reliably control."[155]

Proposed Institutional Models and Their Limitations

Analysts have proposed various institutional responses to the diverse security risks of AI, often drawing on analogies from other fields, including arms control and nonproliferation.[156] These models include multilateral frameworks for inspection, verification, and enforcement; for the control of exports of dual-use technologies; and for early warning and crisis response. While some models hold relevant lessons, many analogies begin to fall apart when one looks more closely at the distinctive security challenges posed by AI.

IAEA-Type Organization

The most frequently cited model for addressing AI's security risks is the IAEA, as noted earlier. In May 2023, OpenAI, the creator of ChatGPT, advocated the establishment of a new international organization based on the IAEA to regulate the pursuit of "superintelligence."[157] Under this proposal, companies or governments that pursue AI capabilities beyond a threshold would "need to be subject to an international authority" empowered by the world community to enforce safety standards, conduct audits, launch inspections, and restrict deployments that endanger security. The UN secretary-general quickly endorsed the idea.[158]

Established in 1957 as an autonomous agency within the UN system, the IAEA was designed to facilitate access to nuclear energy while serving as a nuclear weapons watchdog by ensuring member states' compliance with the NPT.[159] To carry out this mission, it conducts safeguard inspections of civilian nuclear facilities and verifies that fissile and other materials are not diverted to clandestine nuclear weapons programs. The 178-member body reports to the UN Security Council and UN General Assembly. Although the IAEA has had notable failures, particularly regarding North Korea, its overall track record in exposing noncompliance and limiting nuclear proliferation—including thus far by Iran—is impressive.

Since AI's capabilities are similarly dual-use, with the potential for globally destabilizing military applications, the appeal of the IAEA model is obvious. Emulating its nearly universal membership status is also compelling, as it could underpin the legitimacy of the proposed organization's mandate to promote collective security.

Yet the enormous differences between AI and nuclear challenges limit the utility of this governance model.[160] First, the opportunities and dangers AI presents are less straightforward than those posed by nuclear weapons. The latter are difficult to build, require a narrow set of material inputs (including fissile material) that are hard to procure, rely on sophisticated technologies and manufacturing processes, and entail ambitious, large-scale programs difficult for even sovereign governments to conceal. AI, in contrast, comprises a broad suite of general-purpose technologies that, except for the most advanced chips, are easily distributed and wide-ranging in their applications. Furthermore, its development is being driven primarily by private technology platforms with little profit incentive to slow innovation. At a minimum, any regime similar to the IAEA and NPT would need to specify the expectations and legal responsibilities of major technology companies.

Second, whereas nuclear weapons pose a concrete threat to humanity, the existential risk posed to artificial general intelligence remains (for now) theoretical, and there is no equivalent (yet) to the dynamics of mutually assured destruction that eventually induced nuclear restraint and arms control between the United States and the Soviet Union.[161] Rather, there is a pell-mell race, particularly between the United States and China, to dominate this new field, further limiting prospects for negotiating a treaty and setting up an associated universal AI governance agency (at least in the short term).

Non-treaty options offer an alternative route to reduce the potentially destabilizing consequences of AI competition. One avenue would be to negotiate basic codes of conduct that set parameters of responsible behavior for the application of AI in critical domains, such as nuclear weapons or cyberspace. Analogous codes have been proposed for other global issues. Given the difficulties of reopening the Outer Space Treaty, for example, which became effective in 1967, some nations have proposed a code of conduct for outer space activities.[162] The United States, similarly, worked with its Arctic Council partners to draft the Ilulissat Declaration, setting out general principles of conduct in the Arctic.[163]

In parallel, major powers might borrow from arms control regimes by developing CBMs.[164] Even in the absence of treaty-based monitoring and enforcement, CBMs, which are voluntary and can be formal or informal, have stabilized major power relations by reducing ambiguities and mistrust—and associated risks of escalation and miscalculation.[165] During the Cold War, CBMs complemented nuclear arms control negotiations between the United States and Western Europe on the one hand and the Soviet Union on the other. Notable measures included the Moscow-Washington hotline following the Cuban Missile Crisis; voluntary observations and inspections of military exercises; and information exchanges on force deployments, weapons programs, and military budgets.[166]

Military CBMs continue to be used globally and may occur unilaterally, bilaterally, regionally, and multilaterally in several forms.[167] The UN Office for Disarmament Affairs lists five categories of measures that seek to facilitate trust: communication and coordination, observation and verification, military constraints, training and education, and cooperation and integration.[168] For AI, suggested military CBMs beyond dialogues and codes of conduct include promoting shared testing and evaluation standards, sharing information on deployments of AI-enabled systems, and clarifying the expected behavior of autonomous systems, as well as the extent of their autonomy.[169] More general CBMs, including among AI labs, have also been proposed, but these diverge from the traditional model.[170]

Multilateral Export Control Regime

Beyond establishing a new UN agency to serve as an AI watchdog and address dual-use risks, some experts have proposed a multilateral export control regime to control AI's key inputs, including advanced semiconductors.[171] This arrangement would be analogous to existing coalitional schemes for nonproliferation and arms control, particularly the Nuclear Suppliers Group, the Australia Group, the Missile Technology Control Regime, and the Wassenaar Arrangement.

The Nuclear Suppliers Group is a forty-eight-member voluntary regime established in 1975 to prevent proliferation by controlling the export of materials, equipment, and technology required to manufacture nuclear weapons.[172] Members agree to adhere to and implement guidelines for responsible supplier behavior consistent with IAEA safeguards for both nuclear and nuclear-related exports and to exchange relevant information. The Australia Group

is a similar arrangement established in the 1980s to control materials needed to produce biological and chemical weapons. It now has forty-two participating countries (plus the EU), who must adhere to common guidelines and control lists and adopt licensing measures.[173]

The Missile Technology Control Regime is an informal political understanding among thirty-five countries today; it was created in 1987 under G7 auspices to limit the spread of rockets and unmanned aerial vehicles.[174] Members adhere to export control guidelines pertaining to specific equipment, software, and technologies. The Wassenaar Arrangement is a forty-two-nation regime established in 1996 to promote transparency and responsibility in transfers of conventional arms, plus related dual-use technologies, software, and goods.[175] It meets annually to review an agreed-upon regulation list and exchange information on deliveries of nine categories of conventional weapons to nonmember nations.[176]

At first glance, this sounds feasible, particularly as some aspects of advanced AI technology remain highly concentrated. The United States, the UK, the EU, and China dominate the development of significant machine learning systems, and over 90 percent of specialized hardware chips are designed and produced in the United States, China, Japan, South Korea, and Taiwan.[177] In principle, imposing tighter regulation of key inputs for advanced AI models, to ensure that they are available only to actors who meet certain international standards, makes sense.

Any such regime would pose major dilemmas, however. A big one for Western governments is how to create an export control regime that both constrains and restrains China, depriving it of critical technology inputs while somehow also incentivizing its responsible behavior. Consider the export controls that the U.S. Department of Commerce implemented in October 2022, which limit China's "ability to obtain advanced computing chips, develop and maintain supercomputers, and manufacture advanced semiconductors."[178] Updated provisions, added in October 2023, impose even greater restrictions on China's access to advanced chips and semiconductor manufacturing equipment.[179] Because this is a unilateral U.S. approach, however, China can still try to obtain functional equivalents of U.S.-sourced technology from other countries.

To close this gap, the United States is starting to globalize these export controls, including through agreements with Japan and the Netherlands.[180] Some U.S. national security experts have called for making Washington's "small yard, high fence" approach even more multilateral, notionally by updating (or even replacing) the Wassenaar Arrangement to restrict trade in the most advanced chips with stronger enforcement mechanisms.[181] China is not a member of the Wassenaar Arrangement (or of the Australia Group, or the Missile Technology Control Regime)—though Russia is, adding a diplomatic complication. The risk in such a containment approach is that China will seek to undermine global export controls on AI, forming AI alliances with other irresponsible players.

The export control model presents two other complications. First, while chips and hardware might lend themselves to such an arrangement, it would be harder to enforce limits on advanced AI models, algorithms, and other digital inputs. Second, like the IAEA model, the success of any export control regime would require unprecedented government oversight of major AI industry players, including new tools for monitoring end users, to verify *who* uses *what* models and *how*.

Crisis Preparedness and Response

Given the potential dangers posed by malicious use of AI, a number of analysts have proposed crisis monitoring, early warning, and response mechanisms, plus contingency planning based on institutional analogies from international finance and global health.[182]

Such mechanisms have often emerged in the wake of crises. The Financial Stability Board was created among G20 countries (who remain its members) in April 2009, during the global financial crisis.[183] The board's purpose is to develop and strengthen common standards for major cross-border financial institutions—including with respect to capital, liquidity, and risk management. Among other functions, it works closely with the IMF on early warning exercises related to financial instability.[184] Unlike other global economic institutions like the IMF, World Bank, and World Trade Organization, it remains an informal arrangement, which is housed and funded by the Bank for International Settlements and is reliant on a memorandum of understanding among its members. The HLAB has invoked the Financial Stability Board's "macro-prudential framework" as a promising prototype for a "techno-prudential model" with respect to AI.[185]

A different global crisis, namely the COVID-19 pandemic, underscored the urgent need for a new global health surveillance system capable of rapidly detecting emerging (or reemerging) infectious diseases that could pose pandemic threats, a system accompanied by expanded WHO capabilities to orchestrate a quick response to global public health emergencies.[186] In September 2021, the WHO launched the Hub for Pandemic and Epidemic Intelligence to support collaborative monitoring of disease outbreaks; that hub now has a presence in more than 150 countries.[187] To assist with incident response, the WHO also maintains the Global Outbreak Alert and Response Network, comprising "over 250 technical institutions and networks" around the world.[188]

Developing a similar global crisis-response mechanism for AI may be a more feasible avenue, in the short term, for cooperation on collective security. For now, the OECD tracks broadly defined AI incidents but is working on a direct reporting framework.[189] Soon though, a more robust approach that goes beyond incident reporting will be needed, one that involves the participation of more countries and takes action in response to threats.

One conundrum for this crisis preparation and response model is how to address the question of existential risk. In May 2023, hundreds of scientists and industry leaders released a statement, declaring, "mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."[190] To be sure, expert opinion varies wildly on the credibility and nature of the risk posed by rogue AI, as well as the time frame in which any dangers might emerge. However, even a small probability of catastrophe would presumably warrant an insurance policy.

One potential model for early warning and monitoring of a far-out threat is international cooperation on planetary defense. This refers to the capabilities and systems needed to detect and warn of threats to Earth posed by impacts from near-Earth objects—asteroids and comets that orbit the sun—and if possible, efforts to prevent or mitigate such a low-probability but literally high-impact disaster. According to the U.S. government, it is believed that there are as many as 1,000 near-Earth objects larger than 1 kilometer in size capable of devastating the globe; 95 percent of those have been found, and none is on a trajectory to collide with Earth. Another 25,000 larger than 140 meters are believed to exist, but only 42 percent of them have been identified and tracked—and any one of them could destroy a city.[191]

The United States has been at the forefront of efforts to prepare for this threat, leading multilateral diplomacy, supporting technical information sharing and observations, and organizing scenario planning.[192] One could imagine similar multilateral efforts for the existential risk posed by AI as a part of a broader multilateral crisis-preparedness and response framework.

Conclusion

Rather than a single, tidy, institutional solution to govern AI, the world will likely see the emergence of something less elegant: a regime complex, comprising multiple institutions within and across several functional areas. The messy structure of global AI governance will reflect the distinct functional imperatives of AI regulation, the diversity and incentives of relevant public and private actors, and the absence of a single international political authority with the capacity and legitimacy to orchestrate cooperation across multiple domains.

To promote effective AI governance, the emerging regime complex must advance several imperatives:

- provide the world with authoritative, up-to-date knowledge on the state of AI science;

- facilitate the negotiation of common standards and harmonization of AI regulations;

- advance equitable access and benefit sharing, particularly for LMICs; and

- promote collective security by mitigating dual-use dangers, encouraging arms control, and reducing risks from AI itself.

For each of these functions, experts and officials have invoked a plethora of institutional analogies from other global challenges. Yet the direct relevance of these comparisons varies widely. As a general-purpose technology with geopolitically salient implications, and whose development is primarily occurring within the private sector, AI eludes simple comparisons and analogies. Insights from existing institutional models are instructive, but policymakers must still move forward with designing innovative governance approaches to manage AI's unique opportunities, risks, and cooperation dilemmas. As they proceed, they should expect to confront three main challenges.

First, a generic challenge in any regime complex is the danger of incoherence across functional areas. For AI, the objective of encouraging equitable access stands in tension with the goal of reducing the risk of malevolent use or unintended consequences. However, a more daunting challenge will be preventing fragmentation *within* individual issue areas. Major powers, notably the United States and China, are competing fiercely to dominate AI technologies and applications and shape the principles and rules surrounding their governance. Like-minded Western governments can expect to find themselves repeatedly torn between pursuing AI governance within encompassing UN frameworks versus narrower coalitions of the like-minded. Such dynamics, if replicated by China, could well encourage competitive multilateralism, featuring rival minilateral arrangements among subsets of countries committed to competing values or interests. A potential wild card in this geopolitical game will be the postures on AI governance that LMICs in the Global South adopt.

Second, creating new structures for AI global governance—or even increasing the AI competence of existing institutions like the WHO and IAEA—will take time. In terms of the four core functions identified in this paper, the immediate priority task, and the objective most likely to be achieved in the short term given its relatively technical (as opposed to political) nature, is the first. The world will likely move briskly to create an authoritative scientific body, building on the first report on the state of AI science.

The second core function—creating common standards and harmonizing national (or in the case of the EU, regional) regulations—will inevitably take longer, due to major differences in the political cultures, values, and institutions of the major players. This heterogeneity, particularly differences between open and closed societies, may result in multispeed governance, with some (perhaps Western) countries adopting more demanding standards than those that emerge at the UN level.

The question of how to improve access and benefit sharing for developing nations—the third AI governance function—is likely to be fraught, amid growing alienation of Global South nations from the Global North, as well as the intellectual property concerns of private sector platforms. Nevertheless, the current geopolitical context could provide LMICs with leverage to insist on a more inclusive approach to cooperation on AI. At a time of intense competition pitting the West against China and Russia, wealthy nations perceive a strategic need to demonstrate solidarity with developing nations that have been disinclined to align with Western nations, which many regard as indifferent to their plight.[193]

The most difficult governance hurdle is likely to be forging broad multilateral agreement on collective security measures to prevent malevolent applications of advanced AI capabilities and mitigate related dual-use dangers. Given the impediments to universal, treaty-based approaches, countries may begin with narrower, less formal CBMs, codes of conduct, and voluntary export control arrangements. More encompassing approaches may be easier to achieve for crisis preparedness efforts, including early warning systems against potential existential risks that endanger all humanity.

Finally, a recurrent dilemma spanning each of these functional imperatives will be to determine the appropriate role of the private sector as a subject and object of AI global governance. Technology companies are the driving force behind AI breakthroughs and control the lion's share of global AI capabilities. Their participation and cooperation in governance initiatives is critical, but they are also not subject to the same accountability mechanisms as national governments. The question for public authorities is how best to regulate these private actors in a manner that advances the public good, both domestic and global. Compounding this challenge, leading AI labs are themselves divided by stark philosophical differences. Addressing the role of technology companies and other private sector actors across functional areas will require policymakers to diverge from many existing institutional models and pursue novel approaches to global governance.

About the Authors

Emma Klein is a James C. Gaither Junior Fellow in the Carnegie Endowment for International Peace's Global Order and Institutions Program.

Stewart Patrick is Senior Fellow and Director of the Global Order and Institutions Program at the Carnegie Endowment.

Acknowledgments

The authors are grateful to Carnegie President Mariano-Florentino (Tino) Cuéllar and to their colleagues Jon Bateman, Fiona Brauer, Frances Brown, and Steven Feldstein for constructive and insightful comments on previous drafts of this working paper. They also thank Carnegie's Communications team for their meticulous work throughout the publications process and Carnegie India for hosting the Global Technology Summit, whose discussions informed aspects of this paper. Financial support from the Ford Foundation, Hewlett Foundation, and Packard Foundation helped make this research possible.

Notes

1 Sam Altman, Greg Brockman, and Ilya Sutskever, "Governance of Superintelligence," OpenAI, May 22, 2023, https://openai.com/blog/governance-of-superintelligence.

2 Michelle Nichols, "UN Chief Backs Idea of Global AI Watchdog Like Nuclear Agency," Reuters, June 12, 2023, https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12.

3 Jason Hausenloy and Claire Dennis, "Towards a UN Role in Governing Foundation Artificial Intelligence Models," United Nations University, September 9, 2023, https://unu.edu/cpr/working-paper/towards-un-role-governing-foundation-artificial-intelligence-models; Lewis Ho et al., "Exploring Institutions for Global AI Governance," Google DeepMind, July 11, 2023, https://deepmind.google/discover/blog/exploring-institutions-for-global-ai-governance; Anja Kaspersen and Wendell Wallach, "Envisioning Modalities for AI Governance: A Response From the Artificial Intelligence Equality Initiative to the UN Tech Envoy," Carnegie Council for Ethics in International Affairs, September 29, 2023, https://www.carnegiecouncil.org/media/article/envisioning-modalities-ai-governance-tech-envoy; and Robert Trager et al., "International Governance of Civilian AI: A Jurisdictional Certification Approach," University of Oxford Martin School, August 2023, https://www.oxfordmartin.ox.ac.uk/publications/international-governance-of-civilian-ai-a-jurisdictional-certification-approach.

4 Jonas Tallberg et al., "The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research," *International Studies Review* 25, no. 3 (2023): https://doi.org/10.1093/isr/viad040.

5 United Nations AI Advisory Body, "Interim Report: Governing AI for Humanity," United Nations, December 2023, https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf.

6 Matthijs M. Maas and José Jaime Villalobos, "International AI Institutions: A Literature Review of Models, Examples, and Proposals," SSRN Scholarly Paper, AI Foundations Report 1, September 22, 2023, https://doi.org/10.2139/ssrn.4579773.

7 Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford: Oxford University Press, 2023), https://doi.org/10.1093/oso/9780197649268.001.0001; and Stewart Patrick, "Rules of Order: Assessing the State of Global Governance," Carnegie Endowment for International Peace, September 12, 2023, https://carnegieendowment.org/2023/09/12/rules-of-order-assessing-state-of-global-governance-pub-90517.

8 Mark Scott and Clothilde Goujard, "Digital Great Game: The West's Standoff Against China and Russia," *Politico*, September 8, 2022, https://www.politico.eu/article/itu-global-standard-china-russia-tech.

9 Hausenloy and Dennis, "Towards a UN Role in Governing Foundation Artificial Intelligence Models"; Ho et al., "Exploring Institutions for Global AI Governance"; Kaspersen and Wallach, "Envisioning Modalities for AI Governance"; Maas and Villalobos, "International AI Institutions"; and Rumtin Sepasspour, "A Reality Check and a Way Forward for the Global Governance of Artificial Intelligence," *Bulletin of the Atomic Scientists* 79, no. 5 (September 3, 2023): 304–315, https://doi.org/10.1080/00963402.2023.2245249.

10 Yoshua Bengio et al., "Managing AI Risks in an Era of Rapid Progress" arXiv, November 12, 2023, http://arxiv.org/abs/2310.17688.

11 Ian Bremmer and Mustafa Suleyman, "The AI Power Paradox," *Foreign Affairs*, August 16, 2023, https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox.

12 "OECD AI Principles Overview," Organisation for Economic Co-operation and Development (OECD) AI Policy Observatory, May 2019, https://oecd.ai/en/ai-principles; and "Recommendation on the Ethics of Artificial Intelligence," United Nations Educational, Scientific, and Cultural Organization (UNESCO), 2022, https://unesdoc.unesco.org/ark:/48223/pf0000381137.

13 Will Henshall, "E.U.'s AI Regulation Could Be Softened After Pushback," *TIME*, November 22, 2023, https://time.com/6338602/eu-ai-regulation-foundation-models.

14 Nestor Maslej et al., "Artificial Intelligence Index Report 2023," Stanford University, Institute for Human-Centered AI, April 2023, https://aiindex.stanford.edu/report.

15 Daniel Björkegren, "Artificial Intelligence for the Poor," *Foreign Affairs*, August 9, 2023, https://www.foreignaffairs.com/world/artificial-intelligence-poor; Marissa Mock et al., "AI Can Help to Speed Up Drug Discovery — but Only If We Give It the Right Data," *Nature* 621, no. 7979 (September 2023): 467–470, https://doi.org/10.1038/d41586-023-02896-9; "Explainer: How AI Helps Combat Climate Change," UN News, November 3, 2023, https://news.un.org/en/story/2023/11/1143187; Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond, "Generative AI at Work," National Bureau of Economic Research, April 2023, https://doi.org/10.3386/w31161; Devorah Fischler, "Artificial Intelligence Is Leveling Up the Fight Against Infectious Diseases," Penn Today, July 28, 2023, https://penntoday.upenn.edu/news/cesar-de-la-fuente-artificial-intelligence-leveling-fight-against-infectious-diseases; "Artificial Intelligence and the Futures of Learning," UNESCO, September 2023, https://www.unesco.org/en/digital-education/ai-future-learning; and Ian Klaus and Ben Polsky, "Subnational Practices in AI Policy: A Working Guide," Carnegie Endowment for International Peace, December 12, 2023, https://carnegieendowment.org/2023/12/12/subnational-practices-in-ai-policy-working-guide-pub-91186.

16 This refrain about the Global North's focus on risk and the Global South's focus on opportunity was common at the 2023 Global Technology Summit, cohosted in New Delhi, India, by Carnegie India and the Government of India's Ministry of External Affairs. See "2023 Global Technology Summit," Carnegie India, December 4, 2023, https://carnegieindia.org/2023/12/06/2023-global-technology-summit-event-8178.

17 Dan Hendrycks, Mantas Mazeika, and Thomas Woodside, "An Overview of Catastrophic AI Risks," arXiv, October 9, 2023, http://arxiv.org/abs/2306.12001.

18 Jen Easterly, Scott Schwab, and Cait Conley, "Artificial Intelligence's Threat to Democracy," *Foreign Affairs*, January 3, 2024, https://www.foreignaffairs.com/united-states/artificial-intelligences-threat-democracy.

19 Isabelle Bousquette, "Rise of AI Puts Spotlight on Bias in Algorithms," *Wall Street Journal*, March 9, 2023, https://www.wsj.com/articles/rise-of-ai-puts-spotlight-on-bias-in-algorithms-26ee6cc9.

20 "Regulation Essential to Curb AI for Surveillance, Disinformation: Rights Experts," UN News, June 2, 2023, https://news.un.org/en/story/2023/06/1137302.

21 Hao-Ping Lee et al., "Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks," arXiv, October 11, 2023, http://arxiv.org/abs/2310.07879.

22 "Policy Brief: Generative AI, Jobs, and Policy Response," Global Partnership on Artificial Intelligence, 2023, https://gpai.ai/projects/future-of-work/Future%20of%20Work%20Working%20Group%20Report%20v2%20(November%202023).pdf.

23 "Widening Digital Gap Between Developed, Developing States Threatening to Exclude World's Poorest From Next Industrial Revolution, Speakers Tell Second Committee," UN News, October 6, 2023, https://press.un.org/en/2023/gaef3587.doc.htm.

24 Eric Lipton, "As A.I.-Controlled Killer Drones Become Reality, Nations Debate Limits," *New York Times*, November 21, 2023, https://www.nytimes.com/2023/11/21/us/politics/ai-drones-war-law.html.

25 Janet Egan and Eric Rosenbach, "Biosecurity in the Age of AI: What's the Risk?," Harvard Kennedy School, Belfer Center for Science and International Affairs, November 6, 2023, https://www.belfercenter.org/publication/biosecurity-age-ai-whats-risk; and Melissa Parke, "Preventing AI Nuclear Armageddon," Project Syndicate, November 8, 2023, https://www.project-syndicate.org/commentary/dangers-of-artificial-intelligence-ai-applications-nuclear-weapons-by-melissa-parke-2023-11.

26 Chris Vallance, "AI Could Worsen Cyber-Threats, Report Warns," BBC, October 25, 2023, https://www.bbc.com/news/technology-67221117.

27 Michael Hirsh, "How AI Will Revolutionize Warfare," *Foreign Policy*, April 11, 2023, https://foreignpolicy.com/2023/04/11/ai-arms-race-artificial-intelligence-chatgpt-military-technology; Barry Pavel et al., "AI and Geopolitics: How Might AI Affect the Rise and Fall of Nations?," RAND Corporation, November 3, 2023, https://www.rand.org/pubs/perspectives/PEA3034-1.html; and Jacob Stokes, Alexander Sullivan, and Noah Greene, "U.S.-China Competition and Military AI," Center for a New American Security, July 25, 2023, https://www.cnas.org/publications/reports/u-s-china-competition-and-military-ai.

28 James Johnson, "Rethinking Nuclear Deterrence in the Age of Artificial Intelligence," Modern War Institute, January 28, 2021, https://mwi.westpoint.edu/rethinking-nuclear-deterrence-in-the-age-of-artificial-intelligence.

29 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (New York: Oxford University Press, 2016), https://global.oup.com/academic/product/superintelligence-9780198739838?cc=us&lang=en& and Daniel Deudney and Devanshu Singh, "Bounding Power: The ASI Control Problem, Public Safety, and Republican Constitutionalism," Insomnia Quarterly, March 4, 2024, https://isonomiaquarterly.com/archive/volume-2-issue-1/bounding-power-the-asi-control-problem-public-safety-and-republican-constitutionalism.

30 Marietje Schaake, "The Premature Quest for International AI Cooperation," *Foreign Affairs*, December 21, 2023, https://www.foreignaffairs.com/premature-quest-international-ai-cooperation.

31 Adam Satariano, "E.U. Agrees on Landmark Artificial Intelligence Rules," *New York Times*, December 8, 2023, https://www.nytimes.com/2023/12/08/technology/eu-ai-act-regulation.html; "Artificial Intelligence Act: Deal on Comprehensive Rules for Trustworthy AI," European Parliament News, December 9, 2023, https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai; and Raluca Csernatoni, "Charting the Geopolitics and European Governance of Artificial Intelligence," Carnegie Europe, March 6, 2024, https://carnegieeurope.eu/2024/03/06/charting-geopolitics-and-european-governance-of-artificial-intelligence-pub-91876.

32 Kelvin Chan, "Europe Agreed on World-Leading AI Rules. How Do They Work and Will They Affect People Everywhere?," Quartz, December 11, 2023, https://qz.com/europe-agreed-on-world-leading-ai-rules-how-do-they-wo-1851089463.

33 Will Henshall, "The 3 Most Important AI Policy Milestones This Year," *TIME*, December 15, 2023, https://time.com/6513046/ai-policy-developments-2023; Henshall, "E.U.'s AI Regulation Could Be Softened After Pushback."

34 "EU: Artificial Intelligence Regulation Should Ban Social Scoring," Human Rights Watch, October 9, 2023, https://www.hrw.org/news/2023/10/09/eu-artificial-intelligence-regulation-should-ban-social-scoring; Tate Ryan-Mosley, "AI Isn't Great at Decoding Human Emotions. So Why Are Regulators Targeting the Tech?," MIT Technology Review, August 14, 2023, https://www.technologyreview.com/2023/08/14/1077788/ai-decoding-human-emotions-target-for-regulators; and Gian Volpicelli, "EU Set to Allow Draconian Use of Facial Recognition Tech, Say Lawmakers," *Politico*, January 16, 2024, https://www.politico.eu/article/eu-ai-facial-recognition-tech-act-late-tweaks-attack-civil-rights-key-lawmaker-hahn-warns.

35 Will Henshall, "How China's New AI Rules Could Affect U.S. Companies," *TIME*, September 19, 2023, https://time.com/6314790/china-ai-regulation-us.

36 Matt Sheehan, "China's AI Regulations and How They Get Made," Carnegie Endowment

for International Peace, July 10, 2023, https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117.

37 "Interim Measures for the Management of Generative Artificial Intelligence Services," Cyberspace Administration of China, July 13, 2023, http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.

38 Henshall, "How China's New AI Rules Could Affect U.S. Companies."

39 "Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation That Protects Americans' Rights and Safety," White House, May 4, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety.

40 Michael D. Shear, Cecilia Kang, and David E. Sanger, "Pressured by Biden, A.I. Companies Agree to Guardrails on New Tools," *New York Times*, July 21, 2023, https://www.nytimes.com/2023/07/21/us/politics/ai-regulation-biden.html.

41 "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," White House, October 30, 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence; Cecilia Kang and David E. Sanger, "Biden Issues Executive Order to Create A.I. Safeguards," *New York Times*, October 30, 2023, https://www.nytimes.com/2023/10/30/us/politics/biden-ai-regulation.html; and Hadrien Pouget, "Biden's AI Order Is Much-Needed Assurance for the EU," Carnegie Endowment for International Peace, November 1, 2023, https://carnegieendowment.org/2023/11/01/biden-s-ai-order-is-much-needed-assurance-for-eu-pub-90888.

42 Will Henshall, "Why Biden's AI Executive Order Only Goes So Far," *TIME*, November 1, 2023, https://time.com/6330652/biden-ai-order.

43 Upasana Sharma and Shreya Ramann, "AI for All, Beyond the Global North: India's Opportunity?," Carnegie India, November 27, 2023, https://carnegieindia.org/2023/11/27/ai-for-all-beyond-global-north-india-s-opportunity-pub-91110.

44 "National Strategy for Artificial Intelligence," National Institution for Transforming India, June 2018, https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf; "Proposed Digital India Act," Indian Ministry of Electronics and Information Technology, September 2023, https://www.meity.gov.in/writereaddata/files/DIA_Presentation%2009.03.2023%20Final.pdf; and Amlan Mohanty and Shatakratu Sahu, "India's AI Strategy: Balancing Risk and Opportunity," Carnegie India, February 22, 2024, https://carnegieindia.org/2024/02/22/india-s-ai-strategy-balancing-risk-and-opportunity-pub-91693.

45 Charlie Giattino et al., "Artificial Intelligence," Our World in Data, January 31, 2024, https://ourworldindata.org/artificial-intelligence.

46 "Global AI Governance Initiative," Chinese Ministry of Foreign Affairs, October 20, 2023, https://www.fmprc.gov.cn/mfa_eng/wjdt_665385/2649_665393/202310/t20231020_11164834.html.

47 "G7 Leaders' Statement on the Hiroshima AI Process," Japanese Ministry of Foreign Affairs, October 30, 2023, https://www.mofa.go.jp/ecm/ec/page5e_000076.html.

48 Mariano-Florentino (Tino) Cuéllar, "The UK AI Safety Summit Opened a New Chapter in AI Diplomacy," Carnegie Endowment for International Peace, November 9, 2023, https://carnegieendowment.org/2023/11/09/uk-ai-safety-summit-opened-new-chapter-in-ai-diplomacy-pub-90968; and "AI Safety Summit 2023,"

Foreign, Commonwealth, and Development Office; Department for Science, Innovation, and Technology; and the AI Safety Institute, November 2023, https://www.gov.uk/government/topical-events/ai-safety-summit-2023.

49 United Nations AI Advisory Body, "Interim Report: Governing AI for Humanity."

50 OECD, "AI Principles Overview"; UNESCO, "Recommendation on the Ethics of Artificial Intelligence"; and "Our Work," Global Partnership on Artificial Intelligence, accessed March 7, 2024, https://gpai.ai/

projects.

51 United Nations AI Advisory Body, "Interim Report: Governing AI for Humanity."

52 "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023," UK Government, November 1, 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

53 "The Future of AI Governance: A Conversation with Arati Prabhakar," Carnegie Endowment for International Peace, November 14, 2023, https://carnegieendowment.org/2023/11/14/future-of-ai-governance-conversation-with-arati-prabhakar-event-8195.

54 Karen J. Alter and Kal Raustiala, "The Rise of International Regime Complexity," *Annual Review of Law and Social Science* 14, no. 1 (2018): 329–349, https://doi.org/10.1146/annurev-lawsocsci-101317-030830.

55 Robert O. Keohane and David G. Victor, "The Regime Complex for Climate Change," *Perspectives on Politics* 9, no. 1 (March 2011): 7–23, https://doi.org/10.1017/S1537592710004068; "The Global Health Regime," Council on Foreign Relations, June 19, 2013, https://www.cfr.org/report/global-health-regime; and Joseph S. Nye, "The Regime Complex for Managing Global Cyber Activities," Centre for International Governance Innovation, May 20, 2014, https://www.cigionline.org/publications/regime-complex-managing-global-cyber-activities.

56 Keohane and Victor, "The Regime Complex for Climate Change."

57 Alter and Raustiala, "The Rise of International Regime Complexity."

58 Julia C. Morse and Robert O. Keohane, "Contested Multilateralism," *Review of International Organizations* 9, no. 4 (December 1, 2014): 385–412, https://doi.org/10.1007/s11558-014-9188-2.

59 Stewart Patrick, "Four Contending U.S. Approaches to Multilateralism," Carnegie Endowment for International Peace, January 23, 2023, https://carnegieendowment.org/2023/01/23/four-contending-u.s.-approaches-to-multilateralism-pub-88852; and Stewart Patrick, "Multilateralism à La Carte: The New World of Global Governance," Valdai Club, August 7, 2015, https://valdaiclub.com/a/valdai-papers/valdai_paper_22_multilateralism_la_carte_the_new_world_of_global_governance.

60 Ho et al., "Exploring Institutions for Global AI Governance"; Joseph Bak-Coleman et al., "Create an IPCC-Like Body to Harness Benefits and Combat Harms of Digital Tech," *Nature* 617, no. 7961 (May 2023): 462–464, https://doi.org/10.1038/d41586-023-01606-9; and Mustafa Suleyman et al., "Proposal for an International Panel on Artificial Intelligence (AI) Safety (IPAIS): Summary," Carnegie Endowment for International Peace, October 27, 2023, https://carnegieendowment.org/2023/10/27/proposal-for-international-panel-on-artificial-intelligence-ai-safety-ipais-summary-pub-90862.

61 "'State of the Science' Report to Understand Capabilities and Risks of Frontier AI: Statement by the Chair," UK Department for Science, Innovation, and Technology; UK Foreign, Commonwealth, and Development Office; and UK Prime Minister's Office, November 2, 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-state-of-the-science-2-november/state-of-the-science-report-to-understand-capabilities-and-risks-of-frontier-ai-statement-by-the-chair-2-november-2023.

62 "About — IPCC," Intergovernmental Panel on Climate Change (IPCC), https://www.ipcc.ch/about.

63 "About — IPCC," IPCC; "Preparing Reports," IPCC, https://www.ipcc.ch/about/preparingreports; and "IPCC Meets to Approve the Final Component of the Sixth Assessment Report," IPCC, March 13, 2023, https://www.ipcc.ch/2023/03/13/ipcc-meets-to-approve-ar6-synthesis-report.

64 "About IPBES," Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), March 16, 2017, https://www.ipbes.net/node/40.

65 "About IPBES," IPBES.

66 "About Montreal Protocol," UN Environment Programme (UNEP), October 29, 2018, http://www.unep.org/ozonaction/who-we-are/about-montreal-protocol; and "Assessment Panel Overview," UNEP, https://ozone.unep.org/science/overview.

67 "About IPBES," IPBES.

68 Huw Roberts, "A New International AI Body Is No Panacea," E-International Relations, August 11, 2023, https://www.e-ir.info/2023/08/11/opinion-a-new-international-ai-body-is-no-panacea/; and "Will the

World Ever See Another IPCC-Style Body?," *Nature*, March 1, 2023, https://www.nature.com/articles/d41586-023-00572-6.

69 "Global Warming of 1.5°C," IPCC, October 2018, https://www.ipcc.ch/sr15; and "Special Report on the Ocean and Cryosphere in a Changing Climate," IPCC, September 2019, https://www.ipcc.ch/srocc.

70 Gary Marcus, "Two Models of AI Oversight—and How Things Could Go Deeply Wrong," Communications of the Association for Computing Machinery, June 12, 2023, https://cacm.acm.org/blogs/blog-cacm/273791-two-models-of-ai-oversight-and-how-things-could-go-deeply-wrong/fulltext; and Andrea Miotti, "We Can Prevent AI Disaster Like We Prevented Nuclear Catastrophe," *TIME*, September 15, 2023, https://time.com/6314045/prevent-ai-disaster-nuclear-catastrophe.

71 "About — IPCC," IPCC.

72 United Nations AI Advisory Body, "Interim Report: Governing AI for Humanity."

73 "About — IPCC," IPCC.

74 Stewart Patrick, "Reflecting Sunlight to Reduce Climate Risk: Priorities for Research and International Cooperation," Council on Foreign Relations, April 2022, https://www.cfr.org/report/reflecting-sunlight-reduce-climate-risk; and "Biosafety Clearing-House," Convention on Biological Diversity, https://bch.cbd.int/en.

75 Tom Simonite, "Canada, France Plan Global Panel to Study the Effects of AI," *Wired*, December 6, 2018, https://www.wired.com/story/canada-france-plan-global-panel-study-ai.

76 "G7 Leaders' Statement on the Hiroshima AI Process," Japanese Ministry of Foreign Affairs; "AI Principles Overview," OECD; "Recommendation on the Ethics of Artificial Intelligence," UNESCO; Japanese Ministry of Foreign Affairs, "G20 AI Principles," June 2019, https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf; "Global AI Governance Initiative," Chinese Ministry of Foreign Affairs; and "AI Safety Summit 2023," Foreign, Commonwealth, and Development Office; Department for Science, Innovation, and Technology; and the AI Safety Institute.

77 Csernatoni, "Charting the Geopolitics and European Governance of Artificial Intelligence."

78 Hausenloy and Dennis, "Towards a UN Role in Governing Foundation Artificial Intelligence Models"; Ho et al., "Exploring Institutions for Global AI Governance"; Maas and Villalobos, "International AI Institutions"; and Trager et al., "International Governance of Civilian AI."

79 "About ICAO," International Civil Aviation Organization (ICAO), https://www.icao.int/about-icao/Pages/default.aspx.

80 "Convention on the International Maritime Organization," International Maritime Organization (IMO), https://www.imo.org/en/About/Conventions/Pages/Convention-on-the-International-Maritime-Organization.aspx; and "Technical Cooperation," IMO, https://www.imo.org/en/OurWork/TechnicalCooperation/Pages/Default.aspx.

81 "ICAO: Frequently Asked Questions," ICAO, https://www.icao.int/about-icao/FAQ/Pages/icao-frequently-asked-questions-faq-5.aspx.

82 "ICAO: Frequently Asked Questions," ICAO.

83 "Member State Audit Scheme," IMO, https://www.imo.org/en/OurWork/MSAS/Pages/Default.aspx.

84 Craig N. Murphy and JoAnne Yates, *Engineering Rules: Global Standard Setting Since 1880* (Baltimore: Johns Hopkins University Press, 2019), https://muse.jhu.edu/pub/1/monograph/book/66187.

85 Stewart M. Patrick, "Remembering the Washington Conference That Brought the World Standard Time," *World Politics Review*, October 7, 2019, https://www.worldpoliticsreview.com/remembering-the-washington-conference-that-brought-the-world-standard-time; "Internet Assigned Numbers Authority," Internet Assigned Numbers Authority (IANA), https://www.iana.org; and "The Basel Committee," Bank for International Settlements, https://www.bis.org/bcbs/index.htm.

86 Nora von Ingersleben-Seip, "Competition and Cooperation in Artificial Intelligence Standard Setting: Explaining Emergent Patterns," *Review of Policy Research* 40, no. 5 (2023): 781–810, https://doi.org/10.1111/ropr.12538.

87 "ISO," International Standards Organization, https://www.iso.org/search.html.

88 "Artificial Intelligence," International Electrotechnical Commission, https://www.iec.ch/ai.

89 "Technical AI Standards," National Institute of Standards and Technology, August 3, 2021, https://www.nist.gov/artificial-intelligence/technical-ai-standards.

90 "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," White House.

91 "Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology," International Atomic Energy Agency, 2022, https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology; "Ethics and Governance of Artificial Intelligence for Health," World Health Organization (WHO), 2021, https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1; and "EASA Artificial Intelligence Roadmap," European Union Aviation Safety Agency, February 7, 2020, https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-10-published.

92 Kenneth W. Abbott and Duncan Snidal, "The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State," in *The Politics of Global Regulation*, ed. Walter Mattli and Ngaire Woods (Princeton, NJ: Princeton University Press, 2009), 44–88, https://doi.org/10.1515/9781400830732.44.

93 "Mission and Impact of the ILO," International Labour Organization (ILO), https://www.ilo.org/global/about-the-ilo/mission-and-objectives/lang--en/index.htm.

94 "Technical Assistance and Training," ILO, https://www.ilo.org/global/standards/applying-and-promoting-international-labour-standards/technical-assistance-and-training/lang--en/index.htm.

95 "How the ILO Works," ILO, https://www.ilo.org/global/about-the-ilo/how-the-ilo-works/lang--en/index.htm.

96 "How the ILO Works," ILO.

97 "Internet Corporation for Assigned Names and Numbers," Internet Corporation for Assigned Names and Numbers, https://www.icann.org.

98 World Commission on Environment and Development, "Report of the World Commission on Environment and Development: Our Common Future," http://www.un-documents.net/our-common-future.pdf; and Charles Cater and David M. Malone, "The Genesis of R2P: Kofi Annan's Intervention Dilemma," in *The Oxford Handbook of the Responsibility to Protect*, ed. Alex J. Bellamy and Tim Dunne (Oxford: Oxford University Press, 2016), 0, https://doi.org/10.1093/oxfordhb/9780198753841.013.7.

99 United Nations AI Advisory Body, "Interim Report: Governing AI for Humanity."

100 Stewart Patrick, "The Universal Declaration of Human Rights at 75: An Unfinished Revolution," Carnegie Endowment for International Peace, December 7, 2023, https://carnegieendowment.org/2023/12/07/universal-declaration-of-human-rights-at-75-unfinished-revolution-pub-91193.

101 The United States was not a member of UNESCO at the time of the recommendation's signing and has not adopted the recommendation since rejoining UNESCO in July 2023. See "Recommendation on the Ethics of Artificial Intelligence," UNESCO.

102 "Subsidies and Countervailing Measures: Notifications," World Trade Organization, https://www.wto.org/english/tratop_e/scm_e/notif_e.htm.

103 "United Nations Activities on Artificial Intelligence," International Telecommunication Union (ITU), 2022, https://s41721.pcdn.co/wp-content/uploads/2021/06/Executive-Summary-2022-Report.pdf.

104 "Recommendation on the Ethics of Artificial Intelligence," UNESCO.

105 "Proliferation Security Initiative," U.S. Department of State, https://www.state.gov/proliferation-security-initiative; "Artemis Accords," National Aeronautics and Space Administration, https://www.nasa.gov/artemis-accords; and "Declaration for the Future of the Internet," U.S. Department of State, https://www.state.gov/declaration-for-the-future-of-the-internet.

106 "FATF," Financial Action Task Force, https://www.fatf-gafi.org/en/home.html.

107 Trager et al., "International Governance of Civilian AI."

108 "Inaugural Address by EAM, Dr. S. Jaishankar at the Global Technology Summit," Indian Ministry of External Affairs, December 5, 2023, https://www.mea.gov.in/Speeches-Statements.htm?dtl/37337/Inaugural_Address_by_EAM_Dr_S_Jaishankar_at_the_Global_Technology_Summit; and "MoS Rajeev Chandrasekhar Addresses Global Technology Summit 2023," YouTube video, 6:52, posted by DD India, December 6, 2023, https://www.youtube.com/watch?v=kDmZqO6nNLU.

109 "Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law," Council of Europe Committee on Artificial Intelligence, December 18, 2023, https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043.

110 Luca Bertuzzi, "EU Prepares to Push Back on Private Sector Carve-Out From International AI Treaty," Euractiv, January 10, 2024, https://www.euractiv.com/section/artificial-intelligence/news/eu-prepares-to-push-back-on-private-sector-carve-out-from-international-ai-treaty.

111 "Universal Periodic Review," UN Human Rights Council, https://www.ohchr.org/en/hr-bodies/upr/upr-home; "The Paris Agreement," UN Framework Convention on Climate Change, https://unfccc.int/process-and-meetings/the-paris-agreement; and "How the ILO Works," ILO.

112 "Population of Global Offline Continues Steady Decline to 2.6 Billion People in 2023," International Telecommunication Union, September 12, 2023, https://www.itu.int:443/en/mediacentre/Pages/PR-2023-09-12-universal-and-meaningful-connectivity-by-2030.aspx.

113 Stewart Patrick, "The Massive Challenge Facing Leaders at the UN Development Summit," Carnegie Endowment for International Peace, September 12, 2023, https://carnegieendowment.org/2023/09/12/massive-challenge-facing-leaders-at-un-development-summit-pub-90533.

114 "Artificial Intelligence and the Futures of Learning," UNESCO; "Digital in Health: Unlocking the Value for Everyone," World Bank, August 19, 2023, https://www.worldbank.org/en/topic/health/publication/digital-in-health-unlocking-the-value-for-everyone; "WMO Leads New Research Project on Early Warning Systems in Mediterranean," World Meteorological Organization, November 21, 2023, https://wmo.int/media/news/wmo-leads-new-research-project-early-warning-systems-mediterranean; "Food and Agriculture Organization," AI for Good, https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/fao; and "Discovery - AI for Biodiversity Archives," AI for Good, https://aiforgood.itu.int/eventcat/discovery-ai-for-biodiversity.

115 This refrain about the Global North's focus on risk and the Global South's focus on opportunity was common at the 2023 Global Technology Summit, co-hosted in New Delhi, India, by Carnegie India and the Indian Ministry of External Affairs. See "2023 Global Technology Summit."

116 "United Nations Activities on Artificial Intelligence," ITU.

117 "United Nations Activities on Artificial Intelligence," ITU.

118 Amy Paul, Craig Jolley, and Aubra Anthony, "Reflecting the Past, Shaping the Future: Making AI Work for International Development," U.S. Agency for International Development, May 2022, https://www.usaid.gov/sites/default/files/2022-05/AI-ML-in-Development.pdf.

119 "Interim Report: Governing AI for Humanity," United Nations AI Advisory Body.

120 "UK Unites With Global Partners to Accelerate Development Using AI," UK Foreign, Commonwealth, and Development Office, November 1, 2023, https://www.gov.uk/government/news/uk-unites-with-global-partners-to-accelerate-development-using-ai.

121 Björkegren, "Artificial Intelligence for the Poor."

122 Chinasa T. Okolo, "AI in the Global South: Opportunities and Challenges Towards More Inclusive Governance," Brookings Institution, November 1, 2023, https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance; Catherine Cheney, "How Artificial Intelligence Can (Eventually) Benefit Poorer Countries," Devex, January 27, 2023, https://www.devex.com/news/sponsored/how-artificial-intelligence-can-eventually-benefit-poorer-countries-104813; and Yasmine Hamdar, Keyzom Ngodup Massally, and Gayan Peiris, "Are Countries Ready for AI? How They

Can Ensure Ethical and Responsible Adoption," UN Development Programme (UNDP), April 25, 2023, https://www.undp.org/blog/are-countries-ready-ai-how-they-can-ensure-ethical-and-responsible-adoption.

123 Calum Handforth, Alena Klatte, and Yasmine Hamdar, "Thinking DEEP to Ensure AI Delivers the Greatest Impact," UNDP, October 4, 2023, https://www.undp.org/blog/thinking-deep-ensure-ai-delivers-greatest-impact.

124 "Global Digital Compact," UN Office of the Secretary-General's Envoy on Technology, https://www.un.org/techenvoy/global-digital-compact.

125 Maas and Villalobos, "International AI Institutions."

126 Ho et al., "Exploring Institutions for Global AI Governance."

127 "Gavi, the Vaccine Alliance," Gavi (the Vaccine Alliance), January 11, 2024, https://www.gavi.org.

128 "Malaria Vaccine: Questions and Answers on Vaccine Supply, Price, and Market Shaping Update," UN Children's Fund (UNICEF), July 2023, https://www.unicef.org/supply/media/18086/file/Malaria-Vaccine-Questions-and-Answers-on-Supply-Price-and-Market-Shaping-July-2023.pdf.

129 "About the Global Fund," The Global Fund to Fight AIDS, Tuberculosis, and Malaria, https://www.theglobalfund.org/en/about-the-global-fund.

130 Celina Schocken, "Overview of the Global Fund to Fight AIDS, Tuberculosis and Malaria," Center for Global Development, https://www.cgdev.org/page/overview-global-fund-fight-aids-tuberculosis-and-malaria.

131 Eoghan Stafford and Robert F. Trager, "The IAEA Solution: Knowledge Sharing to Prevent Dangerous Technology Races," Centre for the Governance of AI, July 1, 2022, https://www.governance.ai/research-paper/knowledge-sharing-to-prevent-dangerous-technology-races.

132 "Treaty on the Non-Proliferation of Nuclear Weapons (NPT)," UN Office for Disarmament Affairs (UNODA), https://disarmament.unoda.org/wmd/nuclear/npt.

133 "Treaty on the Non-Proliferation of Nuclear Weapons (NPT)," UNODA.

134 "Peaceful Uses Initiative (PUI)," International Atomic Energy Agency (IAEA), https://www.iaea.org/services/key-programmes/peaceful-uses-initiative; and "COMPASS – IAEA Comprehensive Capacity-Building Initiative for SSACs and SRAs," IAEA, https://www.iaea.org/topics/assistance-for-states/compass.

135 "IAEA Low Enriched Uranium (LEU) Bank," IAEA, https://www.iaea.org/topics/iaea-low-enriched-uranium-bank.

136 "Global Nuclear Safety and Security Network (GNSSN)," IAEA, https://www.iaea.org/services/networks/global-nuclear-safety-and-security-network.

137 Daniel Zimmer and Johanna Rodehau-Noack, "Today's AI Threat: More Like Nuclear Winter Than Nuclear War," *Bulletin of the Atomic Scientists*, February 11, 2024, https://thebulletin.org/2024/02/todays-ai-threat-more-like-nuclear-winter-than-nuclear-war; and Yasmin Afina and Patricia Lewis, "The Nuclear Governance Model Won't Work for AI," Chatham House, June 28, 2023, https://www.chathamhouse.org/2023/06/nuclear-governance-model-wont-work-ai.

138 Bulelani Jili, "China's Surveillance Ecosystem and the Global Spread of Its Tools," Atlantic Council, October 17, 2022, https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/chinese-surveillance-ecosystem-and-the-global-spread-of-its-tools.

139 Bill Drexel and Hannah Kelley, "Behind China's Plans to Build AI for the World," *Politico*, November 30, 2023, https://www.politico.com/news/magazine/2023/11/30/china-global-ai-plans-00129160.

140 Tshilidzi Marwala, Eleonore Fournier-Tombs, and Serge Stinckwich, "The Use of Synthetic Data to Train AI Models: Opportunities and Risks for Sustainable Development," United Nations University, September 1, 2023, https://collections.unu.edu/eserv/UNU:9216/UNU-TB_1-2023_The-Use-of-Synthetic-Data-to-Train-AI-Models.pdf; and Cheney, "How Artificial Intelligence Can (Eventually) Benefit Poorer Countries."

141 "The Nagoya Protocol on Access and Benefit-Sharing," Convention on Biological Diversity Secretariat of the Convention on Biological Diversity, January 29, 2024, https://www.cbd.int/abs.

142 "The Nagoya Protocol on Access and Benefit-Sharing," Convention on Biological Diversity Secretariat of the Convention on Biological Diversity.

143 "World Climate Research Programme (WCRP)," World Meteorological Organization, February 4, 2023, https://wmo.int/activities/world-climate-research-programme-wcrp.

144 "Global Digital Compact," UN Office of the Secretary-General's Envoy on Technology.

145 Hendrycks, Mazeika, and Woodside, "An Overview of Catastrophic AI Risks."

146 "Convention on Certain Conventional Weapons-Group of Governmental Experts on Lethal Autonomous Weapons Systems," UNODA, 2023, https://meetings.unoda.org/ccw-/convention-on-certain-conventional-weapons-group-of-governmental-experts-on-lethal-autonomous-weapons-systems-2023; and Izumi Nakamitsu, "Keynote Address: Military Applications of AI," YouTube video, 11:54, posted by Carnegie India, December 11, 2023, https://www.youtube.com/watch?v=FlAfQv3rAYw.

147 Steven Feldstein, "AI in War: Can Advanced Military Technologies Be Tamed Before It's Too Late?," *Bulletin of the Atomic Scientists*, January 11, 2024, https://thebulletin.org/2024/01/ai-in-war-can-advanced-military-technologies-be-tamed-before-its-too-late; and Sarah Grand-Clément, "Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain," United Nations Institute for Disarmament Research, November 10, 2023, https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain.

148 Paul Lushenko, "AI and the Future of Warfare: The Troubling Evidence From the U.S. Military," *Bulletin of the Atomic Scientists*, November 29, 2023, https://thebulletin.org/2023/11/ai-and-the-future-of-warfare-the-troubling-evidence-from-the-us-military.

149 Lipton, "As A.I.-Controlled Killer Drones Become Reality, Nations Debate Limits."

150 "First Committee Approves New Resolution on Lethal Autonomous Weapons, as Speaker Warns 'An Algorithm Must Not Be in Full Control of Decisions Involving Killing,'" United Nations Meetings Coverage and Press Releases, November 1, 2023, https://press.un.org/en/2023/gadis3731.doc.htm; and Bureau of Arms Control, Deterrence, and Stability, "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy," U.S. Department of State, November 9, 2023, https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2.

151 Emilia Javorsky and Hamza Chaudhry, "Convergence: Artificial Intelligence and the New and Old Weapons of Mass Destruction," *Bulletin of the Atomic Scientists*, August 18, 2023, https://thebulletin.org/2023/08/convergence-artificial-intelligence-and-the-new-and-old-weapons-of-mass-destruction.

152 Lauren Kahn, "Ground Rules for the Age of AI Warfare," *Foreign Affairs*, June 6, 2023, https://www.foreignaffairs.com/world/ground-rules-age-ai-warfare.

153 Dan Hendrycks, "Natural Selection Favors AIs Over Humans," arXiv, July 18, 2023, http://arxiv.org/abs/2303.16200.

154 Stephen M. Walt, "Does Anyone Still Understand the 'Security Dilemma'?," *Foreign Policy*, July 26, 2022, https://foreignpolicy.com/2022/07/26/misperception-security-dilemma-ir-theory-russia-ukraine.

155 As of March 2024, the letter has more than 33,000 signatures. See "Pause Giant AI Experiments: An Open Letter," Future of Life Institute, March 22, 2023, https://futureoflife.org/open-letter/pause-giant-ai-experiments.

156 Henry A. Kissinger and Graham Allison, "The Path to AI Arms Control," *Foreign Affairs*, October 13, 2023, https://www.foreignaffairs.com/united-states/henry-kissinger-path-artificial-intelligence-arms-control.

157 Altman, Brockman, and Sutskever, "Governance of Superintelligence."

158 Nichols, "UN Chief Backs Idea of Global AI Watchdog Like Nuclear Agency."

159 "International Atomic Energy Agency," IAEA, https://www.iaea.org/about.

160 Ian Stewart, "Why the IAEA Model May Not Be Best for Regulating Artificial Intelligence," *Bulletin of the Atomic Scientists*, June 9, 2023, https://thebulletin.org/2023/06/why-the-iaea-model-may-not-be-best-for-regulating-artificial-intelligence; and Afina and Lewis, "The Nuclear Governance Model Won't Work for AI."

161 Matt O'Brien, "Artificial Intelligence Raises Risk of Extinction, Experts Say in New Warning," Associated Press, May 30, 2023, https://apnews.com/article/artificial-intelligence-risk-of-extinction-ai-54ea8aadc60d1503e5a65878219aad43.

162 Chris Johnson, "Draft International Code of Conduct for Outer Space Activities Fact Sheet," Secure World Foundation, February 2014, https://swfound.org/media/166384/swf_draft_international_code_of_conduct_for_outer_space_activities_fact_sheet_february_2014.pdf.

163 "The Ilulissat Declaration," Arctic Ocean Conference, May 2008, https://arcticportal.org/images/stories/pdf/Ilulissat-declaration.pdf.

164 Ioana Puscas, "AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures," UN Institute for Disarmament Research, October 12, 2023, https://unidir.org/publication/ai-and-international-security-understanding-the-risks-and-paving-the-path-for-confidence-building-measures; and Ge Jun, "A Review of U.S.-USSR Confidence-Building Measures During the Cold War," *China Military Science*, published by Center for Strategic and International Studies (CSIS), February 20, 2016, https://interpret.csis.org/translations/a-review-of-u-s-ussr-confidence-building-measures-during-the-cold-war.

165 "Military Confidence-Building Measures," UNODA, https://disarmament.unoda.org/convarms/military-cbms.

166 Michael Krepon, "Conflict Avoidance, Confidence-Building, and Peacemaking," in *A Handbook of Confidence-Building Measures for Regional Security*, ed. Michael Krepon et al., third edition, (Washington, DC: Stimson Center, 1998), https://www.stimson.org/wp-content/files/CBMHandbook3-1998-krepon.pdf.

167 "Military Confidence-Building Measures," UNODA.

168 "Military Confidence-Building Measures," UNODA.

169 Michael Horowitz and Paul Scharre, "AI and International Stability: Risks and Confidence-Building Measures," Center for a New American Security, January 12, 2021, https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures.

170 Sarah Shoker et al., "Confidence-Building Measures for Artificial Intelligence: Workshop Proceedings," arXiv, August 3, 2023, https://doi.org/10.48550/arXiv.2308.00862.

171 Sujai Shivakumar, Charles Wessner, and Hideki Tomoshige, "Toward a New Multilateral Export Control Regime," Center for Strategic and International Studies, January 10, 2023, https://www.csis.org/analysis/toward-new-multilateral-export-control-regime.

172 "About the NSG," Nuclear Suppliers Group, https://www.nuclearsuppliersgroup.org/index.php/en/about/about-the-nsg.

173 "Introduction — The Australia Group," Australia Group, https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/introduction.html.

174 "MTCR," Missile Technology Control Regime, https://www.mtcr.info/en.

175 "The Wassenaar Arrangement," Wassenaar Arrangement, https://www.wassenaar.org.

176 "List of Dual-Use Goods and Technologies and Munitions List," Wassenaar Arrangement Secretariat, 2023, https://www.wassenaar.org/app/uploads/2023/12/List-of-Dual-Use-Goods-and-Technologies-Munitions-List-2023-1.pdf.

177 Maslej et al., "The AI Index 2023 Annual Report"; and Charlie Giattino et al., "AI Hardware Production, Especially CPUs and GPUs, Is Concentrated in a Few Key Countries," Our World in Data, January 31, 2024, https://ourworldindata.org/artificial-intelligence?insight=ai-hardware-production-especially-cpus-and-gpus-is-concentrated-in-a-few-key-countries#key-insights.

178 "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China," U.S. Department of Commerce, Bureau of Industry and Security, October 7, 2022, https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file.

179 William Alan Reinsch, Matthew Schleich, and Thibault Denamiel, "Insight Into the U.S. Semiconductor Export Controls Update," Center for Strategic and International Studies, October 20, 2023, https://www.csis.org/analysis/insight-us-semiconductor-export-controls-update.

180 Gregory C. Allen, Emily Benson, and Margot Putnam, "Japan and the Netherlands Announce Plans for New Export Controls on Semiconductor Equipment," Center for Strategic and International Studies, April 10, 2023, https://www.csis.org/analysis/japan-and-netherlands-announce-plans-new-export-controls-semiconductor-equipment.

181 Shivakumar, Wessner, and Tomoshige, "Toward a New Multilateral Export Control Regime."

182 Maas and Villalobos, "International AI Institutions."

183 "Members of the FSB," Financial Stability Board (FSB), January 31, 2024, https://www.fsb.org/about/organisation-and-governance/members-of-the-financial-stability-board.

184 "About the FSB," FSB, November 16, 2020, https://www.fsb.org/about.

185 United Nations AI Advisory Body, "Interim Report: Governing AI for Humanity."

186 "The Independent Panel's Recommendations for Transforming the International System for Pandemic Preparedness and Response," Independent Panel for Pandemic Preparedness and Response, https://recommendations.theindependentpanel.org/main-report/05-the-independent-panel; and "Rapidly Detecting and Responding to Health Emergencies," WHO, https://www.who.int/activities/rapidly-detecting-and-responding-to-health-emergencies.

187 "WHO Hub for Pandemic and Epidemic Intelligence," WHO, 2021, https://cdn.who.int/media/docs/default-source/2021-dha-docs/who_hub.pdf?sfvrsn=8dc28ab6_5.

188 "Global Outbreak Alert and Response Network," Global Outbreak Alert and Response Network, https://goarn.who.int.

189 "OECD AI Incidents Monitor," OECD, https://oecd.ai/en/incidents-methodology.

190 O'Brien, "Artificial Intelligence Raises Risk of Extinction, Experts Say in New Warning."

191 "National Preparedness Strategy and Action Plan for Near-Earth Object Hazards and Planetary Defense," National Science and Technology Council, Planetary Defense Interagency Working Group, April 2023, https://www.whitehouse.gov/wp-content/uploads/2023/04/2023-NSTC-National-Preparedness-Strategy-and-Action-Plan-for-Near-Earth-Object-Hazards-and-Planetary-Defense.pdf.

192 "National Preparedness Strategy and Action Plan for Near-Earth Object Hazards and Planetary Defense," National Science and Technology Council, Planetary Defense Interagency Working Group.

193 Matias Spektor, "In Defense of the Fence Sitters," *Foreign Affairs*, April 18, 2023, https://www.foreignaffairs.com/world/global-south-defense-fence-sitters.

Carnegie Endowment for International Peace

The Carnegie Endowment for International Peace is a unique global network of policy research centers around the world. Our mission, dating back more than a century, is to advance peace through analysis and development of fresh policy ideas and direct engagement and collaboration with decisionmakers in government, business, and civil society. Working together, our centers bring the inestimable benefit of multiple national viewpoints to bilateral, regional, and global issues.

Global Order and Institutions Program

The rules-based world order is under unprecedented strain, buffeted by geopolitical competition, populist nationalism, technological innovation, transnational threats, and a planetary ecological emergency. Carnegie's Global Order and Institutions Program analyzes the shifting landscape of international cooperation and identifies promising new multilateral initiatives and institutions to advance a more peaceful, prosperous, just, and sustainable world.

CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

CarnegieEndowment.org