Global Governance: Goals and Lessons for AI

2024

Global Governance: Goals and Lessons for AI

2024

Contents

Acknowledgements

Thank you to the following experts, who contributed their time and knowledge to this report.

Christina Parajon Skinner
Assistant Professor
The Wharton School, University of Pennsylvania

Sir Christopher Llewellyn Smith
Emeritus Professor of Theoretical Physics - University of Oxford
Former Director General of CERN (1994-1998)

David Heffernan
Chair, Transportation & Trade Practice Group
Cozen O'Connor

Diana Liverman
Regents Professor of Geography
University of Arizona, Tucson

Julia C. Morse
Assistant Professor of Political Science
University of California, Santa Barbara

Rachel Schwartz
Associate, Cozen O'Connor

Trevor Findlay
Principal Fellow
University of Melbourne

Youba Sokona
Vice Chair
Intergovernmental Panel on Climate Change

Foreword

As AI policy conversations expanded last year, they started to be punctuated by unexpected abbreviations. Not the usual short names for new AI models or machine learning jargon, but acronyms for the different international institutions that today govern civil aviation, nuclear power, and global capital flows. ICAO, IAEA, FATF, and FSB were in the mix, alongside IPCC and CERN, two institutions that facilitate critical scientific research across borders.

This piqued our curiosity. We wanted to learn more about how approaches to governing civil aviation might apply to a set of digital technologies that would never be assembled in a hangar or guided by air traffic control officers. And we were eager to learn about nuclear commitments that emerged in an entirely different geopolitical era to regulate technology that showed promise as a tool but had only been used as a weapon.

Our curiosity set us on a journey to learn more about international analogies for AI governance. Through research and referrals from colleagues, we identified a global group of experts who had studied relevant international institutions or participated in them directly. We focused on a range of institutions, including: the International Civil Aviation Organization (ICAO), the European Organization for Nuclear Research (CERN), the International Atomic Energy Agency (IAEA), the Intergovernmental Panel on Climate Change (IPCC), the Financial Action Task Force (FATF), and the Financial Stability Board (FSB).

In October, we had the pleasure of hosting the group at our Redmond campus for a day-long workshop. In a wide-ranging discussion that traversed history, politics, economics, and the law, we covered the missions, functions, and evolutions of these institutions, highlighting the lessons they offer for the global governance of AI. We came away with a rich set of insights and lots of follow-up questions that we subsequently dug in on.

This publication pulls together the product of our learning journey so far: an institutional case study or governance theory chapter from each of our experts, as well as our own reflections on directions for AI governance at the global level. We offer it as a resource to share our learnings with the broader AI policy community

and to spur further reflection and discussion about goals and lessons for governing AI globally.

A key takeaway for us has been that our question should be less about which institutional analogy is most apt for global AI governance and more about the multiple governance functions that apply to AI.

Through this lens, each institution we studied has relevance for international AI governance. Defining global standards, as ICAO does; driving international scientific consensus, as IPCC does; and managing emergent global stability risks, as the FSB does, are all important functions for AI.

As we recognized the relevance of multiple institutional functions, we sought to zoom out and put them in a wider governance context. We defined three desired outcomes of international AI governance:

1. **Globally significant risk governance:** We must manage globally significant safety and security risks that affect us all and on which there's broadly shared agreement regarding the need for coordinated action, such as AI-powered acceleration of chemical or biological weapons development or the deployment of increasingly autonomous systems.

2. **Regulatory interoperability:** We must build international frameworks that help to facilitate and strengthen the coherence and interoperability of domestic policies and regulations across borders.

3. **Inclusive progress:** We must ensure broad access to AI's benefits, fostered through an inclusive global community that contributes to AI research, development, and deployment.

It became apparent that the governance functions we distilled from our analysis of existing international institutions could help secure multiple international AI governance outcomes. For instance, defining and facilitating consistent implementation of standards or codes of conduct is a governance function pursued by ICAO, IAEA, FATF, and the FSB. Having common standards and codes of conduct, in turn, is a key enabler of globally significant risk governance and regulatory interoperability.

This web of functions and outcomes especially matters for the current historic moment. When ICAO was formed as World War II came to a close, formal, treaty-based commitments were more likely to gain traction. Today, "regime

complexes" of formal and informal international organizations "coordinating and competing over policy space" define 21st century global governance.[1] A web of institutions and initiatives, pursuing overlapping and intersecting functions and outcomes, will continue to play key roles in AI governance.

Key desired international AI governance outcomes

1. Globally significant risk governance

International collaboration to monitor for and respond to globally significant **safety and security risks**

2. Regulatory interoperability

International framework to facilitate and strengthen the **interoperability** of domestic policies and regulation

3. Inclusive progress

International network to broaden **access to infrastructure and skilling** for inclusive AI research and development and technology benefits

To get the most traction out of this international AI governance system, we need common frameworks and clear areas of focus to track our progress toward shared goals. We need clarity on where we are today in pursuing these shared goals and where there are gaps that will benefit from coordinated investment and further thinking.

Since we hosted our workshop last October, governments have made tremendous progress. The Hiroshima AI Process defined an *International Code of Conduct for Developers of Advanced AI Systems* (Code of Conduct); the United

Nations General Assembly voiced support for many elements of the Code of Conduct; and the Organization for Economic Co-operation and Development (OECD) initiated a process to develop a mechanism to monitor the application of the Hiroshima Code of Conduct by organizations that choose to adopt it. The UK hosted the inaugural AI Safety Summit, passing the baton to the Republic of Korea ahead of the series moving to France with its AI Action Summit in February 2025. Vital work to advance scientific understanding of AI safety has progressed, with countries around the world agreeing to develop an international network of AI Safety Institutes[2] and the publication of the interim International Scientific Report on the Safety of Advanced AI.[3]

But we are still in the early days of our AI governance project. To achieve the international AI governance outcomes that we've offered here, more work is required, including on developing common frameworks that will act as durable guides for an evolving system. What follows is our further reflections on those frameworks, leveraging what we learned through dialogues with the experts whose insights are captured in case studies, as well as our ideas for concrete next steps to advance further along the path towards those outcomes.

As we continue to develop, implement, and continuously improve AI guardrails, we remain committed to learning about and contributing ideas on AI governance. Most of all, we are excited about what effective governance of our emerging AI economy[4] will mean for people, organizations, and our shared humanity. History tells us that, if we get governance right, a powerful new technology could fundamentally improve countless lives around the world—in ways we can anticipate today and ways that we may later look to with wonder.

Brad Smith
Vice Chair and President

Natasha Crampton
Chief Responsible AI Officer

1

Frameworks and Outcomes for International AI Governance

Few leaps forward in technology and policy innovation compare to what the world has recently experienced. Artificial intelligence (AI) models have proliferated, their capabilities rapidly progressing.[5] Global policy has likewise developed apace, with AI's promise and peril animating discussions in cities ranging from Brussels to DC, Delhi to London and Paris, Santiago to Tokyo and Verona, and many places in between. One thing has become clear: There is widespread determination to act, both to govern how AI is developed and deployed and to apply recent lessons about technology's power as a tool and a weapon.

But act how, where, and toward what more specific outcomes? These questions are more perplexing than they may appear at first glance, and parallels between technology and policy innovation continue to be instructive in understanding our progress with them.

If 2023 was the year of exploration and framing, then 2024 is shaping up to be the year where many new efforts are brought to ground as we further understand the practical application and implementation of technology and policy frameworks. Users are asking more tactical and operational questions about when and how they can put AI technologies to work. Likewise, developers and implementers of AI policy are testing how higher-level objectives can be realized in practice.

International AI governance discussions fit this pattern as well. Last year was one of high-level institutional analogies, with the roles of the International Civil Aviation Organization (ICAO), the European Organization for Nuclear Research (CERN), the International Atomic Energy Agency (IAEA), the Intergovernmental Panel on Climate Change (IPCC), the Financial Action Task Force (FATF), and the Financial Stability Board (FSB) all referenced in the context of global AI governance needs. This year, among other international organizations and initiatives, the United Nations (UN) High-Level Advisory Body (HLAB) on AI, a group to which Microsoft's Chief Responsible AI Officer contributes in her personal capacity,

has considered key questions surrounding the opportunities and enablers of AI, the risks and challenges of AI, and the international governance of AI, including the governance functions needed and the institutional arrangements for carrying them out.[6]

Bringing higher-level policy frameworks to ground requires sorting out where, how, and by whom various objectives are pursued. While international institutions are a key part of pursuing safe, secure, and trustworthy AI, other actors have important and complementary roles in realizing those objectives as well.

Applying learnings from other domains, and in particular civil aviation, nuclear power, and global capital flows, three interrelated layers of AI governance are needed: industry standards, domestic regulation, and international governance. For AI, this is how we see these three layers fitting together:

- First, **industry standards** and specifications for AI safety, security, and trust support policy implementation, bringing together state-of-the-art practices and guardrails based on operational learnings. Industry contributions are important because, like many other 21st century technologies, AI is being pioneered by the private sector. Civil society and academia also influence specifications and standards, contributing research and insights on how proposed measures or controls achieve objectives.

- Second, **domestic regulation** builds on consensus-based standards and practices. Domestic policies and regulation may be focused on a specific sector, domain, issue, or layer of the AI economy (e.g., health, privacy, provenance, or AI applications)—or they may be more horizontal and comprehensive, taking on a broader set of interests, risks, desired outcomes, and AI economy actors.

- Third, **international governance** also builds from consensus-based standards or specifications and complements domestic regulation. Bilateral or multilateral agreements and policies play a complementary role to global policies pursued by international organizations. International governance takes on issues that particularly demand or benefit from cross-border collaboration, including safety or security imperatives or opportunities to facilitate global innovation and economic development.

Key AI technology and policy moments

December 2022

ChatGPT reaches over 1 million users in less than 1 week

February 2023

Release of Bing Chat

March 2023

Release of GPT-4

May 2023

Japan initiates *Hiroshima AI Process (HAIP)* at G7

June 2023

Hugging Face adds 100,000 AI models since January

July 2023

Release of Llama 2; US organizes voluntary commitments from AI companies

August 2023

China implements *Interim Measures for the Management of Generative AI* Services

October 2023

Release of DALL-E 3; Chile and UNESCO host *Ministerial on the Ethics of AI;* US releases AI Executive Order; G7 agrees to HAIP Code of Conduct

November 2023

UK hosts AI Safety Summit; release of ChatGPT Plus and Microsoft 365 Copilot

December 2023

EU agrees on the *AI Act;* India hosts *Global Partnership on AI (GPAI)* Summit; UN High-Level Advisory Body on AI releases interim report

January 2024

Swiss Call for Trust and Transparency in AI Action 1 launches

March 2024

UN adopts AI resolution, *"Seizing the opportunities of safe, secure and trustworthy AI systems for sustainable development"*

May 2024

Countries around the world agree to develop an international network of AI Safety Institutes alongside AI Seoul Summit

June 2024

G7 Leaders commit to stepping up AI governance interoperability efforts and welcome an OECD pilot reporting framework for the *HAIP Code of Conduct*

July 2024

GPAI and OECD unite to advance coordination on trustworthy AI

Recognizing these three overlapping layers allows us to home in on the distinct and complementary roles of international agreements and institutions as part of a broader governance structure. It allows us to consider the issues that particularly demand or benefit from cross-border collaboration, driving a need for international AI governance.

What AI governance outcomes are critical at the international level?

Like many modern-day scientific, industrial, and commercial breakthroughs that came before it, AI is the product of cross-border collaboration that it also stands to strengthen. "Top-tier AI researchers" live, work, and collaborate across regions, with the flow of "elite AI researchers" working in different countries expanding between 2019 and 2022.[7] The new AI economy is also international; AI systems are often built with components sourced from different countries and then, via the global connectivity offered by the internet, made available to customers around the world.

Global interconnection underpins AI governance opportunities. Across borders, we share a common interest in defining safety and security rules that are impermeable.[8] People and organizations around the world benefit from accessing the best AI technologies and components without significant technical or compliance barriers. We also stand to benefit both nationally and across humanity if consistent norms and guardrails help accelerate responsible innovation that hastens sustainability and healthcare solutions.

The cross-border nature of AI technology also challenges governance. National-level technical or compliance barriers may develop for a variety of reasons, including value differences that are difficult to reconcile and more minor discrepancies that are nonetheless burdensome to coordinate. In addition, as with other technologies, AI risks transcend borders; an AI system developed in one country could be misused by someone based elsewhere to cause harm in a third country—or even in multiple countries simultaneously, for example, via cyberattack. Aligning and consistently enforcing rules is critical to managing such risks effectively.

Key international AI governance outcomes should be defined in response to these opportunities and challenges and how international governance fits into a

more holistic AI governance framework. From our vantage point, three high-level outcomes are important to pursue at the international level:

1. **Globally significant risk governance,** focusing on the most severe safety and security risks that affect us all and on which there's broadly shared agreement regarding the need for coordinated action, such as AI powered acceleration of chemical or biological weapons development or the deployment of increasingly autonomous systems in safety-critical scenarios;

2. **Regulatory interoperability,** leveraging international frameworks that help to facilitate and strengthen the interoperability of domestic policies and regulation; and

3. **Inclusive progress,** ensuring broad access to AI's benefits, fostered through an inclusive global community that contributes to AI research, development, and deployment.

Achieving these international AI governance outcomes will require progress across a mix of efforts, including bilateral and multilateral agreements, and existing and new processes and institutions. To secure this future, we need clarity as to the core set of enabling functions that will make it possible.

Which international AI governance functions are necessary, drawing upon lessons from the past?

Just as global interconnection both underpins and complicates international AI governance, so too have similar opportunities and challenges existed in other domains. In the decades after World War II came to a close, greater international interconnection boosted research, invention, and commerce, helping to reduce global poverty and enrich many lives. But it also accelerated the spread of weapons and amplified cross-border criminal activity, leading to safe havens for bad actors and facilitating their access to enabling resources. Governments responded by collaborating to define shared expectations, enforce rules, and share resources.

AI is in many ways unique, and the task of further developing an international governance system for a technology that will continue to rapidly evolve is formidable—but history holds many lessons. In contemplating the international governance functions AI compels, there are useful parallels with institutions and systems created during the post-World War II period to address scientific,

industrial, and commercial breakthroughs. This includes ICAO, CERN, IAEA, IPCC, FATF, and the Bank for International Settlements (BIS), which hosts the Basel Committee for Banking Supervision and the FSB.

Key desired international AI governance outcomes

1. Globally significant risk governance

International collaboration to monitor for and respond to globally significant **safety and security risks**

2. Regulatory interoperability

International framework to facilitate and strengthen the **interoperability** of domestic policies and regulation

3. Inclusive progress

International network to broaden **access to infrastructure and skilling** for inclusive AI research and development and technology benefits

As the relevance of these institutions for international AI governance has been referenced over the past year, at Microsoft, we've sought to learn more from experts who have studied or participated in them directly. We invited these experts to campus for a workshop discussion, during which we sought to more deeply understand why the institutions were created and what their impact has been—as well as contemplate broader international governance trends. To help share our learnings with the broader AI policy community, we invited each expert to submit an institutional case study or governance theory chapter.

This publication brings together these submissions, which we see as offering context and analogies for international AI governance. Dr. Julia Morse provides a historical overview and analysis of international governance, which is followed by five case studies on:

- ICAO, authored by David Heffernan and Rachel Schwartz;
- CERN, authored by Professor Sir Christopher Llewellyn Smith;
- IAEA, authored by Dr. Trevor Findlay;
- IPCC, authored by Diana Liverman and Youba Sokona; and
- FATF, BIS, Basel, and the FSB, authored by Christina Parajon Skinner.

The workshop discussion and expert submissions helped distill for us that our question should be less about *which* institution is most apt for international AI governance and rather more about how *multiple governance functions and institutional purposes* might be relevant to AI and our key desired international AI governance outcomes.

International AI governance functions

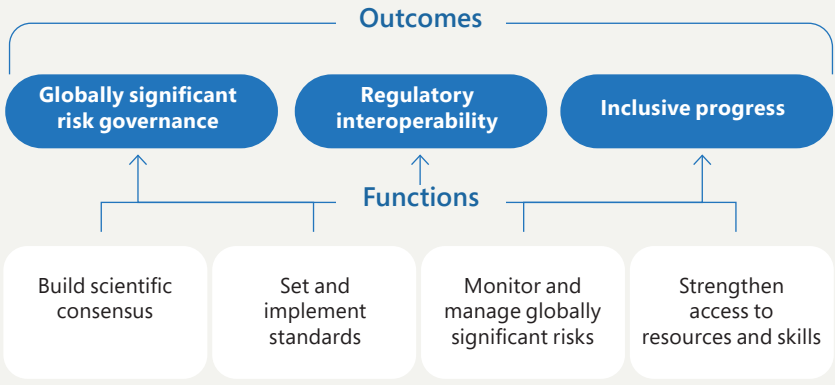
From the authors of each institutional or domain area case study, we learned about the core governance functions pursued in each context. We defined four functions that are not only represented in what each institution was designed or evolved to pursue but also consistent with governance needs that exist for AI. We also recognized how each of these international AI governance functions act as enablers for our desired international AI governance outcomes. This section unpacks that analysis, highlighting how the most relevant international institutions from different domains pursued similar functions.

Monitoring for and managing globally significant AI safety and security risks

This function is closely tied to our desired outcome of globally significant risk governance. Even if interoperable domestic regulations, implemented through well-crafted international standards, enable effective management of many AI risks, technologies and threats will continue to evolve, prompting a need for international coordination on emergent risks with global significance. For example, even as global financial regulators defined general risk mitigation standards, as with Basel, and standards for a specific area of risk, as with FATF, the global financial crisis still transpired—prompting a need for the FSB to both facilitate a response and improve its monitoring of and readiness to mitigate emergent risks.

The FSB and IAEA case studies detail two models for managing globally significant risks. The FSB conducts monitoring or early warning work, identifying emerging financial stability risk and publishing research and working papers that urge attention to certain areas; it also drives forward collective problem solving in areas of high concern to the G20. IAEA requires Members to implement nuclear safeguards whereby states declare the types, amounts, and locations of nuclear materials in their possession. IAEA applies several layers of safeguard measures to ensure that state declarations are correct, including inspections, sample analysis, video monitoring, and satellite imagery. Non-conformance with constraint requirements can trigger UN Security Council action.

A framework for progress



Setting and facilitating consistent implementation of common standards and codes of conduct for AI governance

International standards, ranging from technical specifications to sets of practices or control frameworks against which third parties can certify conformance, will be key to regulatory interoperability and globally significant risk governance outcomes. International standards can also help enable inclusive progress by facilitating the interoperability that enables a global community to access and integrate with global technologies and supply chains.

Standards have a long history of formalizing and advancing best practice and providing implementation details for government-led policy, not only for tangible products growing out of the Industrial Revolution but also for digital services of the current era. An ecosystem of international standards forms the backbone of governance in many sectors, effectively addressing global concerns through a consensus-based mechanism to advance a common approach and reduce barriers to trade and market access.

As Dr. Morse raises and the case studies demonstrate, how institutions develop and implement standards varies. Across ICAO, IAEA, Basel, FATF, and FSB, some institutions focus broadly on an entire industry or sector of the economy,[9] such as civil aviation, whereas others address a specific issue, such as money laundering. However, institutions commonly have governance processes whereby areas of practice in which standards are needed may be identified by governments; then, technical experts, in some cases including stakeholders from academia, civil society, and industry, are convened to develop technical standards or more detailed implementation practices.

There are more marked differences in how adherence to these standards is encouraged or enforced. IAEA, the Basel Committee, and FSB encourage adherence to safety and security or financial governance standards through normative expectation setting and reputational non-compliance costs. Alternatively, FATF and ICAO oversee more intensive monitoring and enforcement regimes—though FATF is not legally established by treaty, and ICAO's impact also depends upon bilateral agreements and domestic monitoring and enforcement. FATF's peer review system of cooperative monitoring has proven nimble and effective at advancing standards adoption, especially when coupled with commercial and reputational costs for non-compliance ICAO conducts safety audits but does not have a direct enforcement role; Member States also audit other states' compliance with standards and, importantly, manage any market access restrictions based on a finding of deficient compliance. Outside of the role of international institutions, there are also processes for mutually recognizing conformance with product safety and security standards. Mutual recognition agreements (MRAs) assert that certification of a product in one country that is party to the agreement is sufficient for that product to be sold across other jurisdictions that are party to the agreement. They have proven popular across a range of product areas; the EU, for example, has MRAs in place for machinery, medical devices, and marine equipment.[10]

Building technical understanding and scientific consensus on AI risks and effective safety practices

This function is a key enabler of all three international AI governance outcomes: globally significant risk governance, regulatory interoperability, and inclusive progress. International collaboration to build technical and scientific consensus on foundational issues, such as how to measure AI capabilities and risks, means more effective use of resources, more consistently understood and applied safety practices, and more aligned interpretations of globally significant safety and security risks. These are each key inputs to the development of effective standards and governance frameworks.

As the case study details, the IPCC is an exemplary model for this governance function. It leverages volunteers from the scientific community, largely academics, to develop research reports that are peer reviewed, reflect global consensus, and are policy relevant. It works best when research is directed by UN Framework Convention on Climate Change (UNFCCC) questions, lending greater credibility to direct the broader research agenda of climate scientists and build from their work. The agreement from ten countries and the EU to develop a global network of AI Safety Institutes, and the publication of the interim International Scientific Report on the Safety of Advanced AI, represent positive progress toward a consensus on AI risk.

Strengthening access to resources needed for inclusive AI research and development and technology benefits

Broad and appropriate access to AI technology and skilling resources is foundational to inclusive progress in a healthy global ecosystem as well as an enabler of regulatory interoperability. Global and local innovation are most impactful when paired together, ensuring that local context helps bridge powerful platform technologies and the needs of diverse communities. In addition, the broader the community that's familiar and interacting with AI technology, the broader our thinking and more inclusive our processes will be for defining and implementing responsible practices. We need individuals and organizations all over the world to be working on responsible AI development, deployment, use, and evaluation, and that broad community needs foundational AI skills to contribute to AI safety practices.

CERN and IAEA offer two models for facilitating access to AI technologies and skills. CERN provides shared infrastructure funded by Member States and Associated Members based on recent net national income; it also requires the publication of research findings and welcomes commercial spinoffs. Most CERN Member States are European, and CERN's formation was in part motivated by an intention to build bridges across states recently in conflict.[11] Alternatively, IAEA's membership is global. As part of its "bargain" with states for complying with nuclear safeguards, IAEA provides technical assistance to support use of nuclear power, as funded by contributions from Member States according to their Gross Domestic Product (GDP).[12]

International institution purposes

In defining the governance functions described above, we applied an AI lens to understand what ICAO, CERN, IAEA, IPCC, FATF, BIS, and the FSB set out or evolved to pursue. However, there's another layer of depth to unpack with regard to the purposes that these and other international governance institutions have historically served.

As further described by Dr. Morse, political scientist Robert Keohane has theorized that there are three purposes for international institutions: facilitating the flow of information; intensifying the consequences of rule breaking; and lowering the costs of cooperation. These purposes cut across a much broader array of international institutions than those highlighted above, surfacing the foundational challenges that international institutions consistently address.

This is an instructive layer to add to our international AI governance framework because it allows us to more directly ask: what kind of problem or opportunity do we need a new or evolving international institution or system of institutions to help address?

- Is there a need to help resolve uncertainty by facilitating information flow;
- Is there a collective action problem that would benefit from more consequences for rule breaking; or
- Are there high transaction costs that necessitate easing or lowering the costs of cooperation?[13]

Imagining pursuit of our international AI governance functions, we can anticipate such challenges or opportunities. For example, if we want to build scientific consensus, we can imagine the need to resolve uncertainty about how the scientific community will prioritize research questions for which there's the most pressing policy need for consensus—or the need to structure a process that reduces the potentially high transaction costs of coordination across a broad global community. The IPCC case study provides experts' perspectives on how this prioritization and coordination works in practice.

Ultimately, each of Keohane's purposes for international institutions overlaps conceptually with the governance functions introduced above and pursued by the institutions from different domains. Studying this overlap helps to illuminate the range of challenges and opportunities that sit beneath each governance function and that motivate the creation or evolution of international institutions.

Studying this overlap also helps to draw out the connections among the governance functions themselves. For example, building scientific consensus may seem a relatively discrete function, with only IPCC being a clear candidate for being dedicated to that function. However, the function could also be considered an enabler of not only every desired international AI governance outcome but also embedded in other functions. Common scientific understanding could support the development and implementation of common standards by facilitating the flow of information underpinning them and lowering the costs of consensus building.

As an analytical tool, Keohane's framework also helps surface two different paths toward a coherent governance system that benefits from these reinforcing functions and purposes. In one path, more common earlier in our post-World War II era, individual institutions may evolve to operate multiple distinct functions, growing their expertise and influence in addressing international cooperation challenges; IAEA epitomizes this approach. Or, in another path, more common later in our post-World War II era, interconnected functions and purposes might be pursued by a system of more and less formally coordinating institutions that help enable and complement each other; the array of institutions that contribute to governance of our global financial system epitomize this approach.

This context also underlines the need for a networked web of institutions and initiatives to work well together, leveraging common frameworks and orientating around key governance functions and outcomes. Durable frameworks for understanding the foundational purposes that international institutions have

played can help direct more coordinated investments in the complementary and reinforcing functions and outcomes needed.

	Strengthening cross-border access to resources or assistance needed for inclusive R&D or benefits **CERN • IAEA**	Building technical understanding of and/or scientific consensus on research key to cross-border issues **IPCC**	Setting and implementing standards and contributing to consequences for non-compliance **ICAO • IAEA Basel • FSB • FATF**	Monitoring for and managing globally significant safety, security, or stability risks **IAEA • FSB**
Facilitating the flow of information	✓	✓	✓	✓
Intensifying the consequences of rule breaking			✓	✓
Lowering the costs of cooperation	✓	✓	✓	✓

Toward a durable and effective consensus on AI risk

To develop effective frameworks for addressing AI risk, we must first build a shared scientific understanding of what the risks are, how they manifest, and how to measure the impact of guardrails. Over the past 12 months, early progress has been made toward these goals. At the November 2023 UK AI Safety Summit, governments agreed to support the development of a "state of the science" report on the capabilities and risks of advanced AI.[14] Through a process chaired by Professor Yoshua Bengio and informed by an expert advisory panel with representatives from 30 nations as well as the EU and UN, an interim International Scientific Report on the Safety of Advanced AI was issued in May 2024, the first global scientific study of AI risk.[15] Then, drafts of the UN Global Digital Compact, including most recently in July, proposed establishing an International Scientific Panel on AI under the auspices of the UN, with the aim of drawing upon existing research and identifying knowledge gaps through a regular cadence of international reports.[16]

This progress is important for two reasons. First, it has helped build momentum around a crucial governance function—building scientific consensus—that will support progress on safety standards, regulatory interoperability, and effective governance. Second, it demonstrates how the efforts of different initiatives and institutions, with their own competencies and histories, can be combined in pursuit of a foundational governance need.

As countries seek a more institutionalized mechanism for scientific consensus building, they should reflect on what competencies are needed and how newly developed capabilities can be built atop existing institutional foundations. As Dr. Julia Morse outlines in Chapter 2, today's frameworks for global and multi-lateral governance are interwoven networks of institutions and initiatives, or "regime complexes," that have evolved over time. They combine newer organizations, typically less formalized and nimbler, with the post-World War II institutional order that Morse describes as an ongoing "foundation" for international cooperation. This mix of old and new, more and less formal, is a key theme across domains for which international governance systems have more recently emerged.

The UN offers an important foundation, complementary to newer initiatives, in AI governance conversations. In many ways, its development in the mid-1940s marked the dawn of a new era in global governance; its evolution charts the history of modern international collaboration. Expanding from 50 countries in the wake of World War II to 193 today, the UN plays many key roles, including as a convenor. The UN brings together diverse perspectives, skills, and interests across some of the issues most important to society today, from conflict prevention to sustainable development to climate change. Already, the UN is driving progress on AI. The UN resolutions on AI governance and capacity building—both adopted in 2024—exemplify the way in which it can strengthen consensus on key priorities. Its work to build global capacity and apply AI to the sustainable development goals will be central to humanity's efforts to address our major challenges.

As Diana Liverman and Youba Sokona set out in Chapter 3, the UN also offers lessons on how to build scientific consensus. The principles that underpin the IPCC reports—that they should be scientifically credible, advance consensus, and be policy relevant but not policy prescriptive—should also guide work on AI. Developing an effective process for consensus building on AI risk will also require countries to address some of the new and specific challenges AI poses on this front, stemming from the speed of its progress and the evolving interplay between AI and complex, real-world systems. As countries make progress toward mechanisms

for building scientific understanding and consensus, they should reflect both on the lessons of processes like the IPCC and the new competencies needed. This will likely center on the need to:

1. **Develop new scientific understanding and keep pace with technological progress.** Whereas IPCC was designed to draw upon existing as well as direct new investments in research on the science of climate change, significant advancements in foundational research are needed to elucidate the capabilities and risks of advanced AI. In addition, whereas IPCC can largely rely on academic researchers, foundational research on advanced AI requires input from a diverse range of experts, including those in industry working at the cutting edge of enhancing AI capabilities and testing for and mitigating risks. Processes must also be responsive to the rapidly developing nature of AI, facilitating quick progress and frequent updates—making the inclusion of those with direct knowledge of the development and deployment of advanced systems and insights into the technology's near-term trajectory even more important. Creating mechanisms to share industry knowledge in a way that preserves the scientific credibility of a reporting process will be key. AI Safety Institutes (AISIs), as centers of technical and research expertise, can play an important role given their deep study of advanced systems and their connection to AI companies. Their involvement could also create an important feedback loop between a report setting out globally significant risks and AISIs' work on risk evaluation and mitigation.

2. **Advance consensus and shared reference points.** Alongside helping to drive the generation of new knowledge, a process for developing and delivering a recurring report must also build this new knowledge into a shared global consensus and advance common points of reference. The UN, with its ability to convene broadly and inclusively, is well suited to a role on this front. While the details of how this role might be realized are important, a multistakeholder UN process could facilitate scientific input and help distill a consensus view relevant for policy discussions. Collaboration with multidisciplinary experts and existing scientific networks will be key to leveraging the appropriate reference points to orient analysis. Just as coordination on climate change is oriented around a shared understanding of the carbon atom as a key unit of measurement, organizations like the OECD and the network of AISIs could help build consensus on benchmarks and metrics that can function as shared reference points for measuring AI risks of concern.

3. **Facilitate an inclusive and durable process.** To build a credible and centralized evidence base over time, the process of developing and delivering a recurring report must be put on secure footing. Likewise, to sustain credibility, the process of building and assessing the rigor of scientific findings must be inclusive. To this end, an established organization, such as the UN, could potentially backstop a secretariat with deep experience facilitating multistakeholder input on technical and policy issues. The role of a secretariat should focus on coordination, overseeing a streamlined process with a clear remit to produce a regular cadence of reports informed by the realities of the technology. The structure of the process for the recent interim Report, with an expert chair overseeing a process that includes soliciting input from multidisciplinary experts with relevant expertise, could serve as an important starting point and be bolstered by inputs from industry, civil society, and scientific networks, including the network of AISIs.

As with recent international governance efforts in other domains, a combination of institutional mechanisms will likely be needed for a durable and effective process for building scientific consensus. As countries move to institutionalize the important work around the recent interim report, they should think about the competencies needed to advance scientific understanding and consensus over the longer term and the different organizations that can contribute to an inclusive and durable process.

Toward international AI governance outcomes in 2024 and beyond

This first half of this chapter has set forth frameworks to put international AI governance efforts in context. It has offered a high-level framework for AI governance, recognizing complementary roles for industry standards, domestic regulation, and international governance. It has then overlaid an international AI governance framework, proposing desired outcomes and functions and weaving in political science theory on international institutional purposes. Working in concert, these frameworks provide breadth and depth to a perspective on why and how we are collectively acting.

And acting we are. Leveraging the desired international AI governance outcomes we've defined, this final section reflects on recent progress, challenges,

and opportunities. It proposes next steps and offers ideas about where energy might be directed in the longer term.

Globally significant risk governance

The world is closer to the start of its AI journey than the end. Given the impressive innovation we have seen over the last 18 months, it is easy to forget that AI is a set of relatively new technologies. In the same way that other general-purpose technologies like the printing press, electricity, and the combustion engine have gone through many iterations, the bulk of AI development and innovation is likely still ahead of us.

Increasingly capable AI will offer significant opportunity, accelerating scientific discovery and addressing major challenges; it may also pose increased safety and security risks. A bad actor might intentionally misuse powerful AI tools as weapons to develop a new pathogen or perpetrate a cyberattack. As more capable models are applied even more broadly across society, the risk of significant accidental damage may also increase. Using AI to help manage critical infrastructure, for example, could pose significant harm if not equipped with safety brakes and operated by appropriately trained individuals.

Some of the most serious safety and security risks of highly capable AI will transcend and manifest across borders.[17] As with other domains presenting globally significant safety, security, or stability risks, such as aviation and financial services, a framework for addressing these risks must therefore be international. Below, we set out ideas about how to build upon efforts already in motion to develop a framework for managing globally significant safety and security risks grounded in the following areas of action:

1. Developing international safety and security standards through a global network of AI safety institutes and partners
2. Requiring notification of highly capable AI model development and advancing an international agreement for government-to-government information sharing
3. Licensing compute providers to validate their operation of secure infrastructure and verify developers meet international safety and security standards as applied by their home government

Developing international safety and security standards through a global network of AI Safety Institutes and partners

International collaboration on safety research is essential if countries are to properly understand and mitigate the risks of advanced AI. Following the development of Safety Institutes in the US and UK late last year, significant progress has been made in institutionalizing safety research and deeper international collaboration. Ten countries around the world have announced Safety Institutes and the EU has further developed its AI Office. This year's AI Summit in Korea witnessed a landmark moment as these countries came together to launch an international network of Safety Institutes to "forge a common understanding of AI safety and align their work on research, standards and testing."[18]

The AI Safety Institutes offer significant promise, serving as hubs for technical safety expertise that can advance scientific understanding and facilitate collaboration across borders. Alongside initiatives like the International Scientific Report on the Safety of Advanced AI, they can help build common understanding and expectations around what risks are most globally significant and ripe for international governance. Risks that AI could facilitate the creation of chemical, biological, radiological, or nuclear (CBRN) or cyber weapons, for example, have been of concern to many, along with the importance of ensuring AI remains under human control.

The AISIs should accelerate work on how to evaluate AI capabilities and risks, sharpening understanding of which models warrant stricter safeguards. This would complement, and potentially, over time, replace compute-based thresholds that can be a useful trigger for ex ante requirements, where necessary, but are a blunt and imperfect proxy for risk and will require revision as algorithmic efficiency improves.

Over time, the network of AISIs and partners could develop a consistent set of evaluation frameworks and metrics, informed by their work to progress scientific understanding. These frameworks could be used by companies as part of internal AI governance or applied by AISIs or other third-party organizations as part of evaluating advanced AI systems where appropriate. Collaboration to develop risk mitigations is also needed. Building on the model of ICAO's technical panels, AISIs could work closely with experts in civil society, academia, and industry to develop consensus best practice mitigations, informed by a deep understanding of the technology and its potential risks. Initiatives like the US AISI Consortium offer a

promising model for ensuring work is informed by deep and diverse expertise from across society.

As with IPCC and UNFCCC, the network of AISIs is likely well positioned to contribute to the research questions on which a broader community of experts offers a consensus view and to bring to bear technical expertise in developing such a report. AISIs across the network could also develop complementary expertise. For example, one AISI could build deep understanding of how to evaluate cutting-edge robotics or agentic systems while another could advance understanding of how to test advanced AI systems for dangerous capabilities.

To make progress on substance, the AISIs should also deepen mechanisms for collaboration. The inaugural convening of AISIs, anticipated in November 2024, should mark the beginning of a regular cadence of meetings, with the potential for dedicated working groups on key lines of effort as well as annual or biannual convenings. Leveraging the AI summit series initiated by the UK in 2023 and carried forward by the Korean and French governments could help reinforce momentum and alignment at the researcher and practitioner levels while also ensuring that the development of technical best practice keeps pace with policymaking.

To ensure the AISI network can draw on scientific and technical talent globally, it will also make sense to further grow the network. Regional AISIs, bringing together experts across a region like Latin America, for example, could help aggregate resources across countries. They could also provide a channel for local experts to input into the AISI network's development of global safety practices and ensure they are fit-for-purpose in a range of cultural contexts.

Requiring notification of highly capable model development and advancing an international agreement for government-to-government information sharing

Countries collaborating through the AISI and partner network could also implement a domestic notification regime for highly capable model training[19] and agree to share high-level information about where such models are being developed in their jurisdictions, helping advance understanding of, and visibility into, emerging risks.

To effectively govern highly capable AI, governments need visibility into where it is being developed and used. Just as aircraft must be registered with domestic authorities, when models hit a high capability threshold—defined through collaboration among the AISI network—developers could be required to provide notification to their home governments, along with information about risk assessment and mitigation measures. Notably, the US and EU governments have taken steps toward this end, with their FLOPs-based thresholds that trigger regulatory reporting.[20]

Governments could leverage the central role that compute providers play in facilitating model training,[21] requiring them to help verify that highly capable model developers provide appropriate notification. Just as anti-money laundering and terrorist financing regulation uses "know your customer" (KYC) requirements to ensure banks track customers engaging in large transactions, KYC obligations could ensure AI compute providers track when model developers use a very large amount of compute above a specified threshold that is indicative of highly capable model training. Moreover, as part of a broader notification framework, AI compute providers could then be required to report to their home government that a model developer is using a very large amount of compute.

Bilateral agreements, such as the MoU between the UK and US to collaborate via AISIs on AI research, standards, and testing, serve as the foundation for broader cooperation and governance in domains beyond AI. For instance, "123 Agreements," which precede significant transfers of nuclear material from the US to partners, also facilitate technical exchanges, scientific research, and safeguards discussions, including via IAEA. Likewise, bilateral agreements are critical to enforcing domestic implementation of safety and security standards defined through ICAO.[23]

Countries collaborating through the AISI and partner network could also work to develop information sharing infrastructure and processes so that they have high-level visibility into highly capable model development globally. One option is to advance an international agreement, with countries committing to collectively requiring that model developers headquartered in their jurisdictions notify them prior to training a highly capable model.

Governments could then share mutually agreed upon information with each other. Recognizing the sensitivities of sharing this type of information, it is likely the information shared between partner countries would be very high level, simply indicating that some level of highly capable model training transpired, and that appropriate notification has been provided. As an additional layer of verification, compute providers could be required to notify their home government that they have granted a customer access to training scale compute following the customer's demonstration of notification.

Governments could then communicate across the framework to match a model developer's training notification with the reporting from the compute provider. Model developers based outside of the agreement would still be able to participate in the framework, but they would be required to notify a government in the agreement that they intend to train a highly capable model. This notification could then be shared across the partner countries involved in the agreement and again verified by providers of training compute. This would help ensure international partners have visibility into circumstances in which entities headquartered elsewhere are training highly capable models using compute providers in framework countries.

This framework could also support a deeper exchange across countries on any emerging challenges with highly capable AI, underpinning the coordinated, rapid response capacity that the FSB provides for global financial stability.

Licensing compute providers to validate their operation of secure infrastructure and verify developers meet international safety and security standards

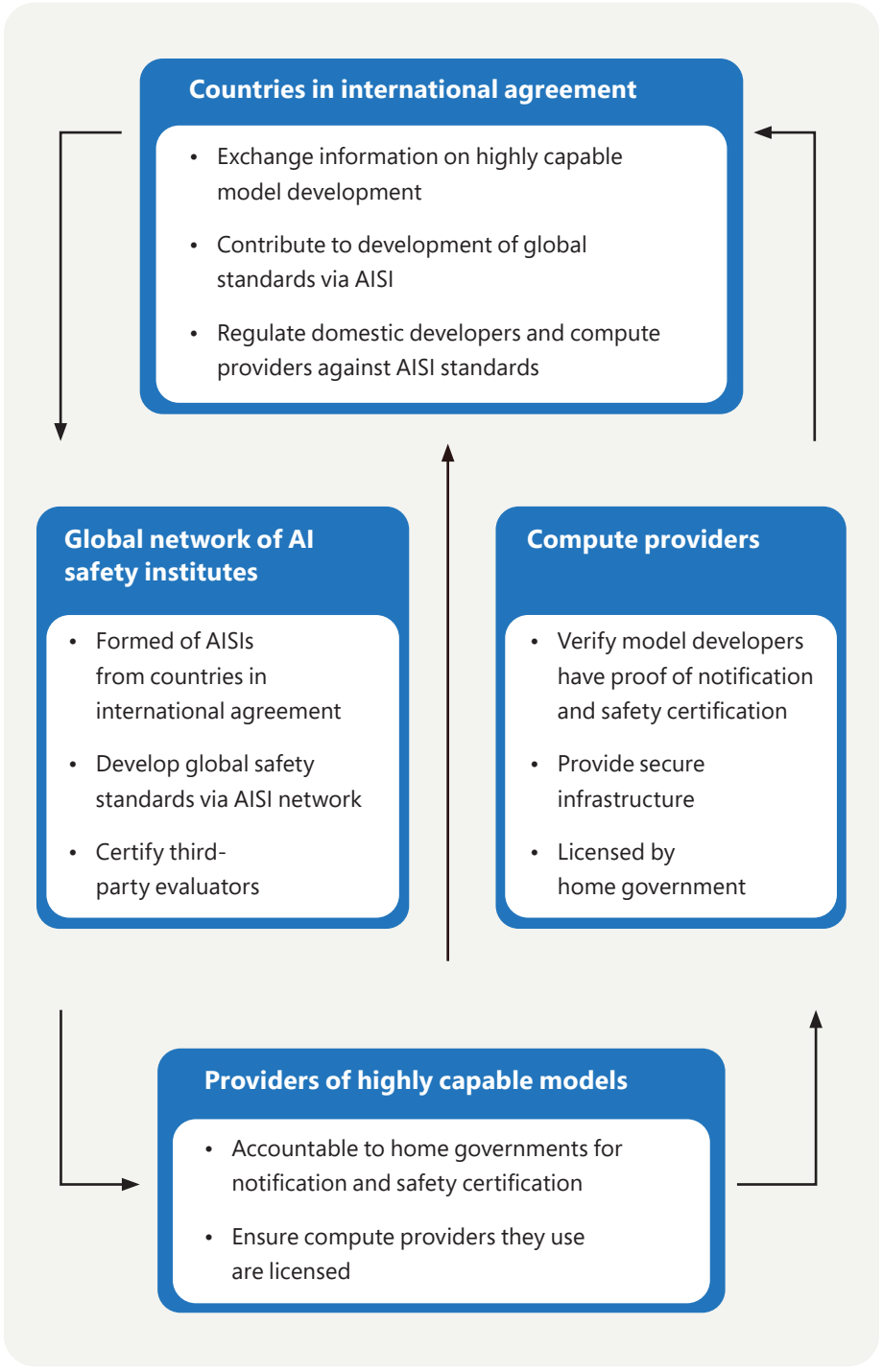
Over time, as scientific understanding and best practice progress and standards for measuring and mitigating globally significant safety and security risks are defined—this framework can act as a foundation to advance additional governance safeguards. Beyond a notification regime, governments could require highly capable model developers or providers in their jurisdiction to apply safety and security measures prior to developing or placing such a model on the market.[22] Highly capable model providers could also be required to undergo safety tests, ensuring that the model does not present CBRN, cyber, or other serious risks.

While safety standards and testing methods would be developed globally via networked AISIs and their partners, those developing and providing highly capable models would be directly accountable to their home government. Approval from a home government or an AISI-accredited third-party assessor could be recognized by others, setting a high bar for safety and security while providing for streamlined regulatory enforcement.

As with the proposed notification framework, compute providers could continue to serve as another important governance node. They could be required to verify that customers training and providing highly capable AI models show they have obtained relevant regulatory approvals from their home government underpinned by consensus safety practices developed by the international network of AISIs and partners. Compute providers could also be obligated to implement international security standards to guard against AI infrastructure being compromised.

Verification and security requirements could serve as core elements of a licensing regime, whereby a license would need to be granted by the compute provider's home government before it's legally authorized to provide compute for highly capable AI model development or hosting.

This framework draws from other models of global governance outlined in later sections. It builds on key concepts from ICAO and FATF, including the way in which globally developed standards are implemented locally in an internationally coherent manner. Ensuring highly capable model developers and providers are subject to direct oversight by their home governments will likely address concerns many countries may have about excessive regulatory inspection risking leakage of sensitive information. As model capabilities continue to improve, it may also play a role in limiting the unintentional proliferation of highly capable models that could be intentionally misused to cause harm. Ultimately, such a framework would build on key efforts already in motion by governments across the globe to advance a durable and effective governance framework capable of responding to emerging and globally significant safety and security risks.


Countries in international agreement
• Exchange information on highly capable model development
• Contribute to development of global standards via AISI
• Regulate domestic developers and compute providers against AISI standards
Global network of AI safety institutes
• Formed of AISIs from countries in international agreement
• Develop global safety standards via AISI network
• Certify third-party evaluators
Compute providers
• Verify model developers have proof of notification and safety certification
• Provide secure infrastructure
• Licensed by home government
Providers of highly capable models
• Accountable to home governments for notification and safety certification
• Ensure compute providers they use are licensed

Regulatory interoperability

While a seamless global framework is especially important for managing significant safety and security risks that governments are concerned about, interoperability among a broader set of policy activities is likewise valuable for global technologies and markets. AI has the ability to help people and organizations around the world achieve more for their businesses, public sector projects, and the planet—but the degree to which it can do so depends on a globally interconnected ecosystem with minimal unnecessary friction. It depends on regulatory interoperability, where there are consistent rules and standards applied to address common expectations for safety, security, rights protection, and trust.

Interoperability has economic, safety, and societal benefits. It enables global organizations to operate more efficiently, directing resources toward rigorous safety and societal risk governance rather than navigation of redundant or inconsistent obligations. It also enables small businesses to access cross-border markets, integrate with global supply chains, and drive innovation. When we say AI has the power to address the world's greatest challenges, we often think of the big breakthroughs underway—but the impact of innovative startups can cascade across industry sectors and parts of the world, catalyzing transformation one developer and one deployer at a time.

Take BeeOdiversity, a Belgian startup. Cofounder Bach Kim Nguyen invented a system that knocks a tiny bit of pollen off worker bees as they return to the hive. Using laboratory analysis and AI models, BeeOdiversity can identify more than 500 pesticides and heavy metals as well as plants. Once they analyze the data, BeeOdiversity scientists make recommendations—including to farmers in Oregon, public water utilities in Europe, and beverage giant AB InBev for its operations in South Africa. In the end, their recommendations not only improve the clients' operations but also the overall environment—and help save bees, which pollinate over 70% of crops that provide the vast majority of food worldwide.[24]

Facilitating and strengthening the interoperability of domestic policies and regulations, which help enable both small and large businesses access markets, grow, and innovate, benefit from three interrelated areas of focus:

1. Defining international principles, policy frameworks, and codes of conduct
2. Supporting consistent implementation through common practices, standards, and expectations for artifacts and evidence
3. Establishing a process to facilitate ongoing collaboration and iterative improvements

Defining international principles, policy frameworks, and codes of conduct

Principles, policy frameworks, and codes of conduct act as a sturdy foundation for interoperable regulation. International principles define common priorities and desired policy outcomes; international policy frameworks define the roles of various stakeholders and areas of potential policy action; and international codes of conduct define sets of more specific common actions that align with areas of focus and accrue to principles.

Iterations of these building blocks have been put in place for AI governance by existing international institutions. In 2019, the Organization for Economic Co-operation and Development (OECD) adopted AI Principles[25] that domestic governments and multilateral organizations, including the G20, have endorsed or leveraged to promote responsible AI.[26] In 2021, UNESCO adopted its *Recommendation on the Ethics of Artificial Intelligence,* endorsed by 193 Member States, providing a framework of values, principles, and areas of action to link higher-level objectives with more practical application approaches.[27]

Last October, the Group of Seven (G7) Hiroshima AI Process (HAIP) agreed to an *International Code of Conduct for Advanced AI Systems* (Code of Conduct), defining a set of actions for responsible AI development and deployment.[28] In March, the UN General Assembly Resolution on *Seizing the opportunities of safe, secure, and trustworthy artificial intelligence systems for sustainable development* (UNGA AI Resolution), adopted by consensus of all UN member states, extended support for the HAIP Code of Conduct approach, broadening a

shared commitment to consistent policies and actions to promote safe, secure, and trustworthy AI.[29]

While the progress made on defining common principles, frameworks, and codes of conduct and growing support for them has been critical, ultimately, realizing their value is dependent upon taking further steps. Clear expectations for how governments and industry can consistently implement interoperable policy frameworks are needed.

Supporting consistent implementation through common practices, standards, and expectations for artifacts and evidence

While leveraging common principles, policy frameworks, and codes of conduct as reference points for domestic AI regulation is a critical step towards interoperability, if jurisdictions adopt high-level actions or provisions but miss opportunities to coordinate further, then they risk creating unnecessary barriers for global commerce and AI safety. As efforts shift toward more detailed implementation, as they already are with the *EU's AI Act and US AI Executive Order*, significant questions can emerge, and divergences in how countries respond to them, even if unintended, can meaningfully disrupt interoperability.

The HAIP Code of Conduct's actions and the UNGA Resolution's provisions offer an up-to-date and focused set of priorities, but they need to be further defined through explanatory guidance and more clearly enumerated expectations regarding implementation. Such guidance and expectations could help direct the efforts of AI developers and deployers and ensure that jurisdictions intending to align to the Code of Conduct or Resolution are interpreting actions and provisions consistently.

The important role of standards in supporting aligned implementation of higher-level policy is well recognized. For example, the EU and US have emphasized a "shared interest in supporting international standardization efforts" in their joint Roadmap for Trustworthy AI and Risk Management.[30] The HAIP Code of Conduct also encourages organizations to "contribute to the development... and use of international technical standards,"[31] and the UNGA AI Resolution stresses the urgency of cooperating on "internationally interoperable safeguards, practices

and standards that promote innovation..."[32] Within the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC), a joint technical committee on AI is developing or has published numerous standards. ISO/IEC 42001, Artificial Intelligence Management System (AIMS), was published earlier this year, providing an international standard that can help organizations implement responsible AI processes and procedures and provide a basis for third party certification.[33] ISO/IEC 42005 will also provide detailed guidelines on implementing and conducting an AI system impact assessment.[34]

While international standards should play an important role in defining implementation details and expectations, other reference points can also be valuable, especially in the near term. Standards for implementing all of the HAIP Code of Conduct actions and UNGA AI Resolution provisions do not yet exist; moreover, the development of standards through a consensus-based process is time intensive. The parallel development of other reference points, such as best practice implementation guidance, not only addresses near-term gaps but also informs potential future standards.

Recent progress with the HAIP offer an especially promising path forward with implementation of the Code of Conduct. In March, the G7 Digital Ministers tasked the OECD to develop a mechanism to monitor the application of the Code of Conduct by organizations that commit to its actions on a voluntary basis,[35] also recognizing a potential role for other stakeholders, such as the Global Partnership on AI (GPAI) and UNESCO. In April, the OECD established an informal task force and kicked off an effort to develop a reporting framework, progress of which was welcomed by G7 leaders in June as they acknowledged a forthcoming pilot.[36] In July, OECD launched the pilot phase of the reporting framework, welcoming organizations to participate by filling out a survey by early September.[37]

A reporting framework that supports organizations in demonstrating implementation of Code of Conduct actions can help to ensure more interoperable regulation globally—especially if such a process is undertaken in advance of or in parallel to domestic regulatory implementation efforts. It can serve as a reference point for domestic efforts, providing greater clarity on key terms, questions and explanatory guidance that bridge from policy objectives to implementation details, best practices distilled from industry reports, and descriptions of potential artifacts or evidence through which organizations can demonstrate implementation, such as templates for documentation. While managing international and domestic efforts proceeding in parallel is complex, as with implementation of the HAIP Code

of Conduct, the EU AI Act, and US AI Executive Order, trying to retrofit domestic policy to align with global approaches is more arduous than leveraging common reference points from the outset.

Establishing a process to facilitate ongoing collaboration and iterative improvements

Fostering regulatory interoperability will be an iterative process. Building common reference points, broadening participation and feedback loops, and considering new ways to support consistent implementation will be critical. Codes of conduct themselves should not be static over prolonged periods of time, especially in areas as dynamic as AI. The process of implementation, especially among a diverse group, is also likely to regularly surface areas for potential elaboration or improvement.

The value of developing international implementation reference points to support domestic regulatory implementation underpins the work now underway to develop the HAIP Code of Conduct reporting framework. Instantiating an iterative pilot this year offers two key advantages. First, it allows the reporting framework to be built out in time to function as a consistent reference point for implementation of the EU AI Act, US AI Executive Order, and other global regulatory activity. Second, it allows for a process whereby governments, the OECD, and partners can surface and discuss challenges and what might be needed to maximize collective investments in a well-regarded reference point.

Broadened collaboration is also needed. With the July announcement that GPAI and OECD are embarking on a new integrated partnership, there's an opportunity to leverage more extensive multistakeholder expertise and to align policy approaches across 44 countries from six continents—and to welcome new members committed to the OECD Recommendation on AI.[38] There's also an opportunity to engage with the HAIP Friends Group, launched on May 2 as an initial set of 49 countries, to broaden the integrated partnership or otherwise support progress on advancing coordinated international efforts on trustworthy AI.[39]

Going forward, the OECD and GPAI could build upon progress thus far on the Code of Conduct reporting framework. They could analyze information provided

by organizations that participate in the Code of Conduct reporting framework process, help distill emerging best practices, and surface gaps that require further consideration or work. Drawing connections and aligning with OECD policies and tools, such as draft due diligence guidance for responsible AI,[40] and with other complementary frameworks, guidelines, and standards that emerge over time could also help demonstrate how the Code of Conduct and reporting framework can function as building blocks for interoperability, linking with other reference points that go deeper on specific issues, such as standards for provenance or safety evaluations. They could also commit to improving the Code of Conduct reporting framework over time as AI technologies, risks, evaluation science, and risk management best practices continue to rapidly evolve. Coordination with related efforts to advance international governance and accountability, such as the AI summit series initiated by the UK at Bletchley Park and carried forward in Seoul and next year in Paris—could also broaden collaboration and build consensus.

Through an iterative and expansive effort, OECD and GPAI could also work with partners, such as the network of AISIs and an international scientific panel, to evolve the Code of Conduct reporting framework to address known gaps. For example, evaluations of AI products are likely to be an important part of a governance regime, including at the domestic and international levels—consistent with the HAIP Code of Conduct, EU AI Act, and US AI Executive Order calls for evaluation of advanced AI models and high-risk systems, as well as the remits of global AISIs. However, AI evaluation science is today unsettled; measurement techniques and instruments are rapidly evolving, and the need for scientifically valid measurement instruments is increasingly in focus. As AISIs and institutions with similar functions, such as the EU AI Office, are expected to work with others, including industry and other experts, to progress the scientific rigor of AI evaluations and the development of effective techniques and instruments, there will likely be value in refining the HAIP Code of Conduct explanatory guidance and expectations regarding artifacts that committed organizations should share to demonstrate implementation.

Approaches for promoting interoperable implementation of consistent policy could also be refined over time. For example, as a version 1.0 Code of Conduct reporting framework is used and demonstrated as valuable, it could also serve to support domestic regulatory efforts toward mutual recognition. As a leading domestic approach for comprehensive AI legislation, the EU AI Act sets an important precedent for such an approach, acknowledging a role for mutual

recognition where "conformity assessment bodies established under the law of a third country meet the applicable requirements of [the Act] and the Union has concluded an agreement to that extent."[41]

Inclusive progress

At the heart of our AI journey is opportunity. But amidst excitement in domains like AI and sustainability[42] or AI4Science[43]—and recent progress towards protecting the Amazon rainforest,[44] improving cancer care and research,[45] or developing new drugs for global infectious diseases[46]—there is the critical question: opportunity for whom? How do we make sure this AI revolution not only enables the transformations we need for our shared futures, but also helps raise everyone up?

We appropriately look to technologies like AI that could help put us on the right course where we've fallen behind on the UN Sustainable Development Goals (SDGs) and otherwise accelerate our progress, but those results are not a given.[47] We must focus on making them happen, investing in AI development and deployment that is inclusive[48] so that AI technologies can most effectively benefit everyone.

We reflect below on the progress that's been made thus far and where we can and need to go, recognizing the role of investments by industry, national governments, and cross-border or global institutions. We consider three areas of focus:

1. Investing in greater access to infrastructure and models
2. Enhancing AI skills by strengthening and amplifying available resources
3. Promoting and facilitating AI for good

Investing in greater access to infrastructure and models

Broad and appropriate access to AI technologies is needed to empower people and organizations around the world to develop and use AI in ways that will serve the public good. Just like other general-purpose technologies in the past, AI is creating a new sector of the economy, with many different technology

components—from chips to datacenters, data, models, tooling, applications, and distribution channels—offering entry points for innovation.[49]

To achieve democratic access to AI, Internet connectivity is essential.[50] Appropriate access to AI infrastructure is likewise critical, particularly for research communities that foster economic growth and public accountability by analyzing the behavior of models and more broadly advancing our understanding of AI.[51] Appropriate access to models is also important for not only researcher but also developer communities that have a greater understanding of their local challenges and ways AI applications may help solve them.

As discussed above in the context of AI governance functions, broadened global access to infrastructure and models would also enable other international AI governance outcomes, including regulatory interoperability. It would accelerate existing efforts to foster globally interoperable approaches to risk evaluation and other required safety practices.

However, strengthening access to AI infrastructure is a formidable challenge. The high cost of compute resources for the training of large scale AI models has been a barrier for many higher education and nonprofit communities. In addition, there's a rising consensus among key stakeholders that, at the frontiers of model capability, careful release strategies may be necessary until marginal safety and security risks are effectively addressed. This need for careful release strategies underscores why inclusive progress needs to be nested within a broader international governance framework.

National governments, including the US, UK, and Canada, are making significant investments to address gaps.[52] Private sector companies are also making investments to support research communities and the broader ecosystem. Microsoft has expanded our AI research grants program;[53] announced investments of over $17.5 billion in new AI and hyperscale cloud infrastructure in Australia, the UK, Europe, and Japan, along with new partnerships with Mistral AI and G42;[54] and committed to our AI Access Principles, including broader programs to promote more innovation and competition than ever before.[55]

Multilateral investments are also needed. One example of potential regional coordination on shared AI infrastructure got underway in September 2023, with the Multilateral Cooperation Center for Development Finance announcing a grant to support a Development Bank of Latin American and the Caribbean project toward creating a network of high performance computing centers for AI

growth, starting in Chile and the Dominican Republic.[56] The UNGA AI Resolution also provides a strong foundation for collaboration, calling upon Member States and inviting other stakeholders to provide assistance to developing countries by enhancing digital infrastructure connectivity; enhancing access to technology that facilitates developing country participation throughout the lifecycle of AI systems; and enabling innovation-based environments to enhance the ability of developing countries to develop technical expertise and capacities and harness data and compute resources.[57]

Enhancing AI skills by strengthening and amplifying available resources

To build with and use AI technologies most effectively, digital and AI skills are critical. As with infrastructure and model access, effectively enhancing AI skills would also have compounding positive effects on broader international AI governance functions and desired outcomes, including by driving inclusive innovation as well as supporting global readiness to implement a more seamless approach to consistent AI guardrails.

But, also as with infrastructure access, the scale of the challenge is substantial. Many different learning paths may be helpful for people and organizations with different starting points with technology and different anticipated scenarios for interacting with AI—across industries, countries, and languages. The demand for baseline digital skills and specific domain areas, like cybersecurity, also is massive and continues to grow.[58]

Existing international institutions and private sector partners are actively working on skilling resources. For example, UNESCO is developing resources,[59] and UNESCO's AI Business Council has also developed a skilling inventory.[60] Last June, Microsoft launched an AI Skills Initiative, through which we have already reached more than 80 million people worldwide.[61] Microsoft has also invested in AI training programs in Australia, the United Kingdom, Germany, and Spain and via partnership with the American Federation of Labor and Congress of Industrial Organizations (AFL-CIO).[62]

The UNGA AI Resolution, in calling upon Member States and inviting other stakeholders to "provide assistance to developing countries...", also underlines

skilling.[63] Specifically, it calls for: increasing digital literacy; capacity building and knowledge sharing related to AI; and providing technical assistance to developing countries related to AI systems.[64]

In addition to recognizing a broad need to increase digital literacy and build AI capacity, we anticipate value in more in-depth technical assistance in support of other international AI governance functions and outcomes, in particular related to managing globally significant safety and security risks. The emerging network of AISIs and partners discussed above offers a mechanism by which technical assistance could be enhanced, including among new AISIs or similarly functioning government structures ramping up capacity. Such a network, coordinating formally or informally, would benefit from strengthened global readiness to not only monitor for risks but also reinforce consistent monitoring of guardrails.

Promoting and facilitating AI for good

Raising up real-world examples of the use of AI to benefit humanity and bringing together multistakeholder research and development efforts using AI to address some of humanity's greatest challenges are critical to realizing the potential of this new technology. Leveraging these examples and efforts, individuals and organizations driving progress can learn from and build on each other's successes, and institutions and effective governance can provide the infrastructure needed to help lower barriers to their cooperation.

For instance, the ITU manages AI for Good, an inclusive UN platform that aims to identify practical application of AI to advance the SDGs and scale those solutions for global impact.[65] Microsoft's AI for Good Lab likewise functions as a research hub, leveraging big data, our cloud technology, and collaboration with our partners to address global challenges.[66]

Recent multilateral efforts underline the importance of ongoing efforts. In October 2023, the HAIP Code of Conduct, building on the White House Voluntary Commitments from Leading AI Companies to Manage the Risks Posed by AI, called upon organizations to "prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education."[67]

The March UNGA AI Resolution also calls upon Member States and invites others to "accelerate the inclusive and positive contribution" of AI to the SDGs.[68] Likewise, in March 2024, the G7 recognized the need for new multistakeholder partnerships to strengthen AI ecosystems in developing countries, including by democratizing compute power and developing open and secure data models.[69]

We see opportunities for globally coordinated investments in AI for good to be more integrated with those towards investments in skills and infrastructure. For instance, Microsoft's AI for Good Lab works at the intersection of AI for good and digital and AI skills, creating AI tools that can help illuminate gaps in broadband availability at a more granular level. The International Computation and AI Network (ICAIN), an effort launched earlier this year at the World Economic Forum (WEF), likewise aims to work at the intersection of AI for good and appropriate access to infrastructure, envisioning pooling expert knowledge and computing resources "to promote the development of interdisciplinary, innovative research and expertise for large-scale AI models that serve society and the achievement of the [SDGs]".[70]

Orienting for what's next

Over recent months, the pace and scope of international AI governance activities have been encouraging but also dizzying. This chapter has offered frameworks to orient those activities in the broader context of AI governance as well as the longer history of international governance. It has proposed three international AI governance outcomes, highlighting how they relate to efforts at the domestic level and among industry as well as how international governance functions relevant across other domains can help enable those outcomes.

As a vast and multifaceted project, international AI governance will continue to involve multiple institutions and processes, building from today's efforts by the UN, G7, G20, OECD, GPAI, and other organizations and initiatives. Together, this AI governance system will fill in gaps but also leverage the international governance infrastructure and more than 400 formal and informal international organizations referenced in Chapter Two, supporting critical progress on our common objectives for AI safety, security, and trust.

Deepening our understanding of some of these institutions, the international governance systems they've helped form, and their purposes and functions will help us further orient toward our desired outcomes and anticipate the challenges and opportunities ahead. The chapters that follow thus elaborate on historical context, conceptual frameworks, and institutional analogies relevant to the international AI governance project on which the world is now embarking. At this critical moment in 2024, when we need to maintain momentum as we shift to the difficult work of deepening, implementing, and iteratively refining, these reflections can help inform our efforts to define common language and frameworks that reinforce a set of common expectations for how, where, and toward what specific AI governance outcomes we are collectively acting—as well as the most valuable next steps we can take toward those outcomes.

Reflecting on initial feedback

After launching version 1.0 of *Global Governance: Goals and Lessons for AI* in May, we set out on a listening tour. Learning from experts in other domains was our original goal with this book, and we were excited to share the expert contributions that shaped our thinking. From the outset, we also wanted to spur conversation with and gather insights from those contributing to AI governance initiatives around the world.

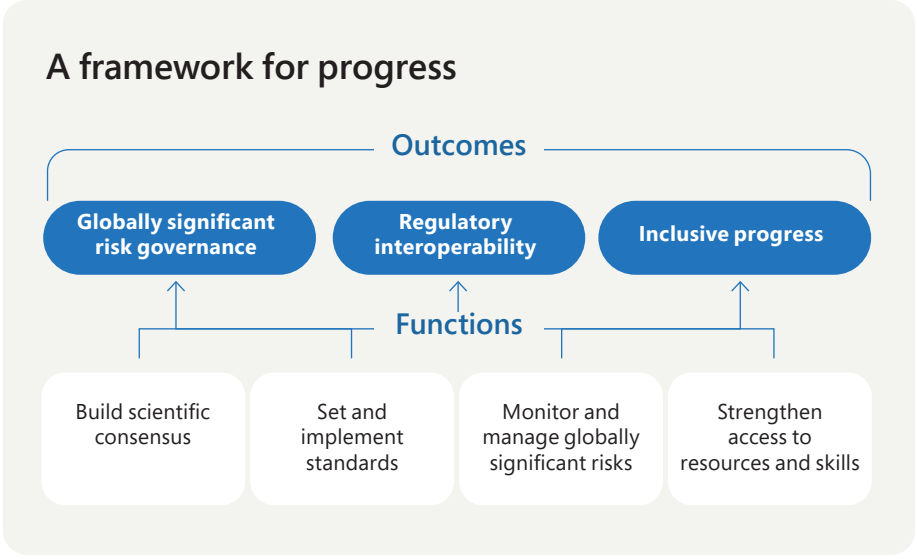
Since publication, we have hosted or joined[71] more than a dozen conversations in which we've presented the book. In May, we held events in Seoul and Tokyo, provided a briefing in Singapore, and joined an event hosted by the Geneva Graduate Institute. In June, we held events in São Luís, Brazil and in Brussels, and we joined events hosted by Eurasia Group in Toronto, the Tony Blair Institute and Centre for Governance of AI in London, Sciences Po and Renaissance Numérique in Paris, and Renaissance Numérique in Brussels. In July, we held events in DC and New York, bringing together US and global diplomats as well as UN officials. In August and September, we held sessions with Responsible AI Fellows in the Global Perspectives program on which we partner with Stimson Center, gathering insights from globally diverse AI experts.[72]

While we continue to solicit feedback, the insights we've heard thus far fall into three broad categories: the shape and contours of AI governance frameworks, the value of international governance analogies, and the broader trends that set the context for advancing international AI governance. We set out some

specifics below, highlighting where participants helped us see opportunities for AI governance conversations to go further.

On the shape and contours of AI governance frameworks

Overall, at a conceptual level, the international AI governance outcomes and functions outlined in Chapter 1 resonated with participants. They recognized the consensus that seems to be emerging on AI governance functions, including through the work of the UN High-Level Advisory Body on AI[73] and Renaissance Numérique.[74] More practically, they were optimistic about OECD initiatives to deepen regulatory interoperability, UN projects to foster inclusive progress,[75] and AI Safety Institute (AISI) collaboration to enable both regulatory interoperability and globally significant risk governance.



Multiple international AI governance functions, which build upon lessons learned from other domains, could help secure multiple international AI governance outcomes.

Participants also offered ideas on how to improve or clarify the scope of the frameworks. Some raised the notion that the "inclusive progress" outcome should more explicitly address equitable development and emphasized the foundational role of capacity building for regulatory interoperability. Progress on climate change was put forward as appropriate for more focus, as was the need

for increased societal engagement in AI governance dialogues. Some participants called for more explicit recognition of whether frameworks are intended to govern AI in civilian or military contexts or both. Others asked how AI governance frameworks can address the risks of fragmentation among sectoral regulation, both domestically and internationally, and examine ove laps with existing technology policy activity related to competition, copyright, privacy, and security.

Among governance functions, many participants were focused on the importance of effective oversight. This makes sense: If an international framework is to effectively address risk and facilitate interoperability in a durable way, then it must grapple with incentivizing and enforcing norms, standards, or rules. However, in AI as in other domains, this often implicates trade-offs with sovereignty. Participants noted that effective oversight has been critical to nuclear power, civil aviation, and financial services governance, with IAEA, ICAO, and FATF differing in approaches to domestic and global enforcement. They observed that domestic and international approaches to government and third-party assurance will be needed but questioned how to effectively pursue both in tandem. They further underscored the complexity of AI assurance and oversight; several participants observed that scientifically valid approaches to assessing and mitigating AI risk are nascent and there are few assessors with sufficient expertise.

For our part, we see the ICAO and FATF models of enforcement distinguished by their domestic enforcement of international standards—as instructive for AI. Recent progress offers building blocks in this direction—a network of AISIs and partners could help define technical safety practices that could then be formalized into international standards and applied by governments. Mechanisms for overseeing the G7 Hiroshima AI Process International Code of Conduct for Advanced AI Systems could help establish consistent expectations for how companies demonstrate implementation of practices, with a draft OECD reporting framework being advanced in pilot form and opportunities to leverage Global Partnership on AI (GPAI) expertise emerging now that GPAI and the OECD have joined forces on AI governance.[76]

On analogies for international governance from other domains

Many dialogue participants also highlighted opportunities to further explore institutional analogies across domains of international governance. They raised limitations with civil aviation and nuclear power analogies for AI due to differences in the pace of fundamental changes in design and application of the technologies being governed. Participants also encouraged us to draw insights from additional domains with interesting parallels and potential lessons for AI, including pharmaceuticals.

In multiple conversations, participants also pointed to the need to analyze global governance failures, impediments, or slow progressions to extract valuable lessons for AI. In some cases, participants offered cybersecurity as an area worth considering given international governance efforts and a perception of mixed results. They raised certifications like Common Criteria, which establishes consistent international standards and evaluation processes, as well as efforts to enforce norms for responsible behavior, such as the UN's ongoing efforts, Microsoft's call for a Digital Geneva Convention, and the Paris Call for Trust and Security in Cyberspace.[77]

We're keen to keep thinking about analogies and distilling lessons from the recent and more distant past. The pace of technology change matters for AI governance, putting different demands on defining and enforcing standards as well as on monitoring for emerging risks. We also take lessons from the civil aviation and nuclear power domains. From ICAO, we see the opportunity to enforce international standards through domestic regulation and bilateral and multilateral agreements. From IAEA, we draw insights on how to effectively bring together multiple important functions, including increasing global access to technical assistance and monitoring for globally significant risk.

On the broader context for advancing international AI governance

Reflecting on frameworks for international AI governance alongside lessons from other domains also helped surface trends that will impact our AI governance path. Many dialogue participants raised the complexity of the emerging AI

governance system. They emphasized that shared frameworks for outcomes and functions are necessary but insufficient; we also need a shared understanding of the roles of various institutions and initiatives in contributing to those outcomes and functions. In many conversations, participants noted the breadth of institutional efforts and initiatives currently underway and the need for a "mapping" or governance "architecture" to help facilitate clarity and collaboration. It has been encouraging to see interest in picking up this work within the "Global AI Governance" working group contributing to the AI Action Summit that France is hosting in February 2025.[78]

As a shared view comes into focus, participants further highlighted that it will be important to institutionalize emerging elements of a global AI governance architecture, including a network of AISIs and partners, an AI summit series, and a state of the science report. They also asked whether and how both "global" and "multilateral" approaches can contribute to international AI governance, especially in what participants observed as a complex geopolitical environment. Building from what we learned from Dr. Julia Morse on the history of international governance, our expectation is that formal and informal institutions and initiatives will need to contribute alongside each other, leveraging complementary strengths and helping to accelerate our collective progress. As is highlighted by Dr. Morse in Chapter 2, formal and informal institutions and processes can compete with each other or they can coordinate, as FATF and the UN have to positive effect.

Participants also consistently called attention to the need for more progress on technical safety to keep pace with AI policy developments. They brought to the fore the risks of a disconnect between technical and policy communities working on AI governance and how meaningful progress depends on better convergence of these conversations. This resonated with our interests in grounding governance in science—including social science and facilitating broader input into technical safety work. It also reinforced our view that collaboration among the initiatives and institutions driving progress on AI governance will be critical. The scientific consensus advanced through a state of the science report, the technical best practices developed by AISIs and partners, and the codes of conduct that deepen governance norms should all work in concert and evolve with one another.

We have appreciated the insights that participants have provided in the events, workshops, and dialogues that we have convened or contributed to thus far. We look forward to ongoing deliberation, ideation, and conversation on these important issues.

2

The Building Blocks of Global Governance: A Comparative Exploration with Lessons for AI

Authored by Julia C. Morse

The modern era is characterized by unprecedented levels of global cooperation. International organizations (IOs) organize state behavior across numerous issue areas, covering everything from high-stakes security concerns like nuclear proliferation and terrorism to complex, technocratic topics like sanitation and food safety. Whereas global governance was once rare and known primarily for idealistic failures like the League of Nations, today more than 300 formal and 150 informal bodies promote cooperation across states. These IOs vary in mandate, membership, and authority, yet each is part of the complex architecture that governs life in the 21st century.

How did we get to this highly institutionalized world? And what lessons do existing IOs hold for incipient AI governance? This chapter links past and present with an eye to the future. Section One begins by recounting the origin story of modern global governance. Despite its cooperative orientation, it was war, not peace, that gave birth to the United Nations and many other well-known IOs. The political tensions of the Cold War spawned additional growth, and the number of treaty-based-IOs more than doubled during this period. Over the last thirty years, however, cooperation has shifted toward more informal bodies, as states seek flexible and adaptive solutions to new types of challenges. As a result of these trends, cooperation has fragmented—even a single-issue area might have ten or more IOs that make relevant policy.

The five domain areas and related IOs examined in this report are best understood within the broader context of this historical trajectory. Each IO is a product of a historical moment when state goals and geopolitical interests aligned

and resulted in a specific mandate, structure, and operations. The objectives and operations of each IO thus offer insights into possible roles for future AI regimes.

To compare and contrast these objectives, Section Two draws on political scientist Robert Keohane's foundational insights into institutionalized cooperation and applies these arguments to the IOs included in this report. First, IOs facilitate the flow of information across cooperating states. They create shared understanding of problems, develop standards for acceptable behavior, and monitor state conformity with the standards. Second, IOs intensify the consequences for rule breaking through reputational mechanisms, external enforcement, and even occasionally institutionalized enforcement. Third, IOs lower the "costs of doing business" so that states and non-state actors can exchange information and develop expertise, provide technical assistance, and even transfer technologies across borders. Comparing the cases along these key dimensions reveals both variation and commonalities across governance models.

Section Three extracts policy lessons from the comparative case analysis. The cases illustrate the importance of strong leadership, particularly from actors with the technical expertise to develop standards and the market power to enforce them. Historically, this leadership has come most often from the United States. The cases also highlight the importance of defining a clear purpose for a new IO. Not all objectives can be accomplished at once, and states may need to make tradeoffs between different goals. Additional lessons highlight how first steps at cooperation may be reinforced over time, as IOs evolve and often strengthen through external processes. Overall, the cases highlight the urgent need to identify common objectives and initiate preliminary governance; many of the fine-grained details will logically follow.

Global governance from 1945 to today

Modern global governance has its roots in war and conflict. Amid the pronounced desperation and fear of the early World War II period, allied countries became convinced that the only hope for establishing a lasting peace lay in the creation of an international organization that would unite countries. In August 1941, US President Franklin D. Roosevelt and UK Prime Minister Winston Churchill laid the foundation for such a body, forging an agreement that affirmed common principles like respect for sovereignty, trade openness, and abandonment of the

use of force.[79] Five months later, twenty-six countries, all at war with the Axis powers, subscribed to these common principles in the "Declaration by United Nations." This was the first time that the term "United Nations" was used, and it stipulated a clear vision for a post-war world.

The next three years saw intense negotiations over the structure and membership of the United Nations, with the US taking a leadership role. Roosevelt wanted to build a strong post-war order where political disputes could be routed through international institutions rather than spilling into military battles. Given the League of Nations' failure to prevent the outbreak of war, Roosevelt was convinced that any new institution needed the power to enforce its decisions and that US involvement was essential. He promoted a framework where core countries like the United States, the Soviet Union, the United Kingdom, and China would provide institutional leadership, and worked to reach compromises that would balance the need for widespread participation with the protection of US interests.[80] The UN's eventual bicameral structure, where enforcement power resides within the 15-member Security Council but budgetary power lies with the inclusive General Assembly, reflects this balance.

The post-war period saw tremendous growth in global governance. The creation of the United Nations in 1945 launched a new trend in which states sought to institutionalize cooperation. In the economic arena, organizations like the World Bank and the International Monetary Fund became key to development and monetary efforts. Security cooperation expanded through regional organizations like the North Atlantic Treaty Organization and oversight bodies like the International Atomic Energy Agency (IAEA). Across issue areas and policy domains, states increasingly turned to IOs. Between 1945 and the end of the Cold War, the number of treaty-based-IOs more than tripled, growing from 66 to 313 in a little more than four decades.[81] The US desire to institutionalize its leadership position, the rise of shared global norms, and the increased number of countries in the global system all likely contributed to this trend.[82]

The post-Cold War period heralded another shift, this time in how states designed new global governance bodies. Formal, treaty-based commitments were poorly suited to address more specialized challenges like combating money laundering, intelligence cooperation after terrorist attacks, and private security during armed conflict. Modern threats required more technocratic and flexible approaches, often with a smaller group of likeminded countries. While legally binding treaties provided stability and policy reassurance, they also took years

to negotiate and involved varied coalitions. States turned to creating task forces, clubs, networks, and forums; informal IOs surged as formal IOs stagnated. Today, there are nearly 150 informal IOs—more than double the number at the end of the Cold War.[83]

Informal global governance is one of the defining features of the 21st century. Such organizations have no legal status, often a small or non-existent secretariat, and fewer members than formal IOs, yet they make decisions with wide-ranging repercussions for states. Informal forums like the G7 and G20 allow states to cooperate and coordinate policy while protecting autonomy. They are also remarkably durable, as states adapt or expand IO missions to address new challenges or increase their authority over time. The Financial Action Task Force (FATF) began in 1989 as a G7 initiative to coordinate anti-money laundering policy, but today the FATF has 39 members plus a vast network of associate countries, and designs standards that cover additional topics like combating the financing of terrorism and proliferation.[84]

Yet while countries turn to informal IOs to address new challenges, the post-war institutional order continues to be the foundation for cooperation. Formal and informal IOs sit alongside each other, coordinating and competing over policy space. These "regime complexes" of multiple IOs that work on a single issue can reinforce each other's actions, as has occurred in the global counter-terrorism arena. States have inserted FATF recommendations on combating terrorist financing into UN Security Council resolutions, lending additional legal clout to "soft law" standards.[85] They may also compete with each other, challenging established rules or international law.[86] As institutions proliferate, the effects of a single IO on policy outcomes become challenging to disentangle from larger patterns of global governance.

Comparative analysis of cases

The five case studies in this report reflect many of the historical trends described above. Treaty-based organizations like the International Civil Aviation Organization (ICAO), CERN, and the IAEA were established in the two decades immediately following World War II, when states viewed multilateral solutions as integral to preventing the outbreak of another war. Indeed, even CERN, an IO centered around research and scientific collaboration, was also intended to foster cooperation between people recently in conflict.

Later cooperative efforts were more varied and encountered different geopolitical challenges. The Intergovernmental Panel on Climate Change (IPCC) was an outgrowth of a formal IO, the United Nations, and has worked to achieve scientific consensus around climate change to support the creation of new legally binding climate change treaties. Yet such efforts have proceeded in fits and starts, as formalized cooperation appears increasingly difficult to achieve in the post Cold War era. Meanwhile, financial governance has expanded significantly over the last fifty years, all the while relying on informal IOs staffed with government bureaucrats.

How can we make sense of such varied institutions with quite different origin stories? Renowned political scientist Robert Keohane theorizes that states create IOs to serve three purposes: facilitating the flow of information, intensifying the consequences of rule breaking, and lowering the costs of cooperation.[87] This theoretical framework sheds light on the achievements and challenges of each global governance example.

Improving information

All IOs exist in part to facilitate the flow of information across states. One common way that information promotes cooperation is when IOs work to build consensus around problem definitions. When states have varied threat perceptions, this task is crucial: how can states work together to solve a problem if they fail to understand it in the same way? IOs can help states define the nature of a challenge, which is often a necessary first step before moving forward with a solution.

Nearly all IOs in this report take on this problem-defining role, but none as important as the IPCC. When scientists and policymakers convened in Toronto, Canada, in 1988 to call for the establishment of an intergovernmental panel on climate change, there was significant uncertainty about the process surrounding global warming, including its attribution to human activities. Each consecutive IPCC report enhanced intergovernmental consensus on the nature of the risk at hand. Because IPCC reports are made public, they also promoted a shared understanding across citizens and private actors.

In addition to defining problems, IOs can help coordinate state expectations around acceptable behavior and best practices. ICAO was explicitly established for this standard-setting purpose: countries needed to develop a single set of

expectations around topics like airspace sovereignty, overfly rules, and air navigation. States also anticipated the challenges that would be posed by differing approaches to airline safety, and the concomitant need to set clear guidelines. Given its influence on industry, ICAO consults heavily with private sector experts when formulating standards, but member states approve the final decisions.

Financial governance institutions are also oriented primarily around improving information, in this case through adaptable standard setting and monitoring. The advantages of such an approach can be seen through the lens of crisis response. The 2008 financial crisis led G20 states to pay renewed attention to topics like financial risk management. In the wake of the crisis, the G20 created the Financial Stability Board and provided existing IOs with core tasks related to enhancing sound regulation in the financial sector and promoting integrity in financial markets.[88] Financial governance institutions responded quickly to this request. Because the Basel Committee's standards are not tied to a specific treaty, finance ministers were able to integrate new information and update the accords, publishing Basel III in 2011. FATF similarly updated standards and intensified its monitoring of state compliance with its standards. FATF's approach, wherein it regularly updates its recommendations and conducts in-depth, peer evaluations of member state policy, is emblematic of the advantages of informal IOs. Without the force of international law, states are more willing to revise standards and subject themselves to intensive monitoring.

Finally, of all the case studies included in this report, the IAEA has perhaps the most important informational role of all: monitoring civilian nuclear programs in an effort to detect diversions for weapons purposes. The IAEA's safeguards regime is one of the most intrusive monitoring regimes in international politics, and it is a product of both its time and the alignment of geopolitical interests on this particular issue. When the IAEA was created in 1957, and when its role shifted to mandatory safeguards with the entry into force of the Nuclear Non-Proliferation Treaty in 1970, treaty based cooperation was the norm and on this rare issue, US and Soviet interests aligned. Moreover, non-nuclear states were told that to gain access to these technologies, they had to submit to the IAEA's procedures, including nuclear material accountancy, on-site inspections, remote video monitoring, and sample analysis. The IAEA's monitoring powers are thus intrinsically tied to the context of this issue: states agreed to a strong, legally binding monitoring regime because they gained access to technologies that otherwise would be unavailable.

Organizing an IO around information provision involves making tradeoffs between different goals. If states are interested in reaching a shared understanding of a threat, then broad participation across both governments and non-state actors will add legitimacy to the effort and make the final outcome more impactful. But this type of widespread information gathering effort may also slow progress on policy action, as it allows countries to deflect cooperation by saying they are waiting for a final consensus. Informal standards, on the other hand, can be established in a timelier fashion, and may incentivize quicker policy action through monitoring. Yet this approach often works best with smaller groups of likeminded states and so policymakers will have to work harder to achieve global legitimacy. Additionally, if states anticipate that countries may be unwilling to follow global standards, a robust and widespread monitoring apparatus will be essential to policy impact.

Intensifying the consequences of rule breaking

A second objective of institutionalized cooperation is to intensify the consequences for rule breakers. International politics has no overarching authority or global policeman, yet the existence of IOs makes it more costly for countries to break agreements or violate established norms. States incur varying degrees of reputation damage for failing to follow through on their commitments. As a result, IO monitoring reports that highlight non-compliance can be a powerful way of incentivizing behavior change.

Both the financial governance institutions and ICAO lean into these reputational mechanisms. Such governance models are built around the assumption that states prefer to have positive reputations in these arenas and will therefore work to modify their behavior to avoid bad publicity. In the case of financial governance, governments want to attract private capital and cross-border investments, and therefore strong financial incentives exist to maintain a positive reputation. In the case of ICAO, governments could face reputational fallout from both citizens and industry if they fall far below international standards.

Reputational mechanisms may reach into the realm of outside enforcement. While most IOs lack formal enforcement powers, states sometimes step in to punish other countries that fail to follow the rules. The ICAO case study provides such an example. The United States and European Union have audit systems based on ICAO standards and may restrict air travel to their jurisdictions if countries

receive poor ratings. Given the size of these economies, such ramifications can be extremely costly for a country's airline industry.

FATF takes this outside enforcement a step farther. Since 2010, the organization has maintained "black" and "grey" lists of countries that are failing to comply with FATF standards. This list is publicized in triannual announcements, and although it is officially not coercive, it has market repercussions. Banks in other countries typically subject clients from listed countries to higher costs and transaction delays, thus imposing direct penalties on the banking sector in listed countries. This market enforcement process has been extremely effective in incentivizing countries to improve their compliance with FATF standards.[89]

Unsurprisingly given the importance of its mandate, the IAEA has the strongest incentive structure to encourage states to follow international rules. If IAEA inspectors detect non-compliance with nuclear safeguards, the IAEA can report a state to the UN Security Council. In February 1993, for example, the IAEA Director General referred North Korea to the UN Security Council after it failed to grant IAEA permission for a special inspection.[90] The Security Council then called upon North Korea to comply with the agreement but refrained from undertaking any significant punitive action until establishing sanctions in 1996.

In cases where some states are likely to ignore international standards or take actions that undermine global cooperation, an IO's ability to create consequences for rule breaking is essential to institutional success. But the optimal system for incentivizing behavior is likely to vary. Reputation can be a powerful mechanism when states share similar priorities, but it may fall short if interests significantly diverge. Outsourcing enforcement to other actors, whether they be states or markets, can be powerful, but it assumes that these actors have clear incentives to punish non-compliant behavior. Finally, creating a strong legal enforcement regime as exists in the IAEA example may be an effective deterrent, but it is unlikely to override core security concerns, particularly when punishment requires widespread agreement among states.

Lowering the costs of cooperation

Finally, IOs may also be designed to lower the costs of ongoing cooperation. Many cooperation problems require ongoing engagement from states. The creation of an IO, particularly one that maintains regularly scheduled meetings where countries are represented by the same delegates year after year, allows

countries to engage with each other in a more efficient manner. Even when the full membership of an IO meets less frequently, IOs typically have subsidiary bodies like ICAO's 36-member Council that are tasked with the more technical aspects of cooperation and adopt procedures for routine meetings and discussions.

Formal IOs may have secretariats that facilitate such processes, providing even basic services like the UN's Blue Book where diplomats can easily find the contact information for their counterparts in other countries. Secretariats may also house experts with specialized knowledge. The IAEA, for example, not only monitors safeguards but also assists developing countries with nuclear technology. Its technical cooperation program provides transfer assistance, helps states identify energy needs, and assists with radiation and nuclear safety. Such ongoing assistance is an important part of the nuclear bargain whereby states are willing to submit to intrusive monitoring.

Lowering the costs of research and scientific collaboration are also common benefits of IOs. CERN has been quite effective in this regard. The existence of a shared space where scientists can converge to focus on a narrow set of topics has led to significant advances in research, and the facility has become a focal point for physicists from all over the world. The IPCC has also ensured ongoing cross-country scientific exchange, both by convening IPCC panels and also by producing rigorously researched reports.

Among informal IOs, ongoing cooperation is typically facilitated through transnational networks of bureaucrats.[91] Financial governance institutions like the Basel Committee are staffed with regulators (typically central bankers and finance officials); FATF meetings are attended by finance officials, central bankers, foreign affairs, and sometimes law enforcement officials. In FATF's case, these bureaucrats are also directly involved in evaluating the policies of peer countries. The meetings and monitoring processes build relationships and make it easier for these officials to engage with each other on relevant policies.

Each IO promotes ongoing cooperation in different ways, and each approach has its own strengths and weaknesses. Concentrating knowledge in a secretariat can build expertise and provide direct points of contact for states seeking technical assistance, yet over time, IO bureaucrats may increasingly expand their authority and operate in ways unanticipated by states.[92] Scientific and research collaboration may promote great leaps forward in understanding and knowledge, yet states have no obligation to integrate such advances into their decision-making or to

act in response to such developments. Finally, bureaucratic networks intensify policy investment in participating states, yet they may operate like clubs that concentrate knowledge in the hands of developed countries and exclude developing economies.

Lessons for AI governance

Understanding the history, politics, and operations of existing global governance regimes illuminates five core lessons for AI. **First, and most essential, any conversations around IO creation must start with establishing clear objectives.** A new IO for AI could create shared expectations around AI risks and develop clear conceptualizations of safety and security, or it could be more action-oriented, focused on standard setting and incentivizing state cooperation. Each approach would necessitate different design choices in terms of membership, governance, and operations. Governments must start by asking themselves: what is the most urgent cooperation problem? If states do not agree that AI poses significant risks and need to build out baseline knowledge before taking additional steps, then perhaps the IPCC model is best. If risks are clear but can be circumvented by establishing best practices, then a standard setting model like Basel might work. Finally, if some risks are clear but states anticipate an unwillingness of some parties to follow established standards, an approach that involves standard setting and outsourced enforcement, such as with FATF or ICAO, might be the best way forward.

Second, political leadership will be paramount to achieving any action in a timely fashion. Stronger IOs require more engagement from politically powerful actors. The policy success of bodies like the IAEA, ICAO, and the FATF is directly linked to support from countries like the United States. The rapid pace of AI developments means that countries need to act quickly, and rapid policy response is most possible when powerful countries are at the forefront of policy action. Notably, leadership on AI governance also has significant strategic advantages, as first movers will have more influence. It is easier to shape incipient norms than to disrupt established ones. Any early IOs in this area will have prolonged effects on the evolution of AI global governance.

Third, while early governance endeavors set the tone for future cooperation, they should not be viewed as final products. Most IOs deepen

their authority and expand their mandates across time. Even formal IOs like the IAEA have adopted new agreements to address gaps in monitoring and enforcement. In the IAEA's case, it has also expanded its governance to include nuclear safety and security. Mandate expansion is particularly common in informal IOs like Basel and FATF. States should not aim to create a full cooperative agreement regulating all aspects of AI, particularly given the rapidly changing nature of the threat. Instead, incremental cooperation may be the best path forward. Focusing on topics that have broad geopolitical consensus, such as preventing the use of AI for the creation of biological weapons, may be one path forward; policymakers may want to delay negotiations on more controversial subjects, such as AI and military technology.

Even amid disagreement about core principles, common ground is still possible if cooperation is oriented around practical applications. Cooperation to combat terrorism is one notable such example. Countries have negotiated 13 international conventions and protocols related to preventing specific types of terrorism, yet no consensus definition exists around the term "terrorism". Indeed, after 9/11, the United Nations Security Council adopted a far-reaching resolution requiring states to take legislative action on terrorism without ever specifying the definition of the term. In contrast to the Council's quick response, countries have been negotiating a comprehensive terrorism convention for more than 20 years through the General Assembly and have yet to reach consensus. If member states had waited for a shared definition of "terrorism," policy action would have been significantly delayed.

Fourth, formal legal authority does not equate to strong policy impact, just as informal status does not mean an IO is ineffectual. States have increasingly turned to informal governance in recent years because it is adaptable and effective in many policy domains. The financial governance institutions in this report have had significant impacts on regulatory policy and the day-to-day practices of global banks. In the FATF case, the organization has diffused its recommendations across 200 economies, despite lacking any legal status. And while this report highlights informal IOs in the financial space, this mode of cooperation is most common in the security realm.[93]

The distinction between formal and informal IOs also does not equate to enforcement. An IO may officially have a strong legal enforcement regime, but the existence of such a mechanism does not mean that states are willing to use it. The UN Security Council has the ability to authorize the use of force—the strongest

possible enforcement of international law that exists in international politics—yet the Council rarely deploys this punishment, even amid significant rule violations. In contrast, both FATF and ICAO have relied on external actors like the private sector and individual governments to enforce compliance with their standards.

Finally, any new cooperative efforts on AI will need to be integrated into the existing global governance infrastructure. More than 400 formal and informal IOs exist today. Within each issue area, a host of different IOs coordinate and compete over policy influence. Even though AI is a new issue area, new IOs will bump up against other policy domains. AI global governance could touch on security, development, climate change, and human rights. Strategic policymakers may be able to leverage the existence of longstanding institutions to reinforce AI governance efforts, using bodies like the Security Council and the General Assembly to endorse new standards.

But to the extent AI governance touches on other policy domains, governments should anticipate calls for inclusion and potential pushback from existing IOs and relevant actors.

Conclusion

The world is at a pivotal moment when it comes to AI. This technology will transform modern society in a myriad of ways, and policymakers have a unique opportunity to shape this transformation. Global governance initiatives are already in incipient stages; now is the time to make crucial decisions about core objectives. IOs are designed to solve specific cooperation problems, and therefore all institutional design proposals should be contingent upon first identifying top priorities. Importantly, global governance can proceed on several fronts at once. It is possible to create one body to assess overall risks, another to set standards and address core security threats, and still another to promote technology transfer. Yet the most urgent priorities are to identify common objectives with likeminded partners and begin to build out a multilateral framework. What starts as a small AI agreement may rapidly expand to become a core feature of 21st century global governance.

3

Institutional Analogies for Governing AI Globally

Building on the comparative exploration offered in the previous chapter, we delve more deeply into the emergence, evolution, and functions of institutions and governance systems that offer analogies and lessons for international AI governance, including:

- The International Civil Aviation Organization (ICAO);
- The European Organization for Nuclear Research (CERN);
- The International Atomic Energy Agency (IAEA);
- The Intergovernmental Panel on Climate Change (IPCC); and
- The Financial Action Task Force (FATF), Basel, and the Financial Stability Board (FSB).

3.1 The International Civil Aviation Organization (ICAO)

Authored by David Heffernan and Rachel Schwartz

Purpose

International commercial air transport is a complex and constantly evolving industry, the success and vitality of which are attributable in significant part to the role of the International Civil Aviation Organization (ICAO), a United Nations (UN) body. The complex and high-stakes nature of safely moving people and goods around the world requires a robust international governance system that provides legal and operational stability and predictability. Since its conception, ICAO has served the civil aviation sector as the industry's global standard-setting agency and facilitator of cooperation among nations in furtherance of a coordinated approach to the fundamental issue of air safety.

History

The Chicago Convention

ICAO is the product of an extraordinary World War II era initiative that led to the signing of the Chicago Convention, an international treaty governing civil aviation. In September 1944, 52 nations represented by over 950 delegates convened in Chicago to negotiate the scope and terms of such a treaty. The conference's purpose was to "make arrangements for the immediate establishment of provisional world air routes and services" and "to set up an interim council to collect, record and study data concerning international aviation and to make recommendations for its improvement."[94] On December 7, 1944, the Chicago Convention was signed and opened for ratification Member States. Today, 193 nations are Member States of the Convention.[95]

The Chicago Convention specifically envisioned an immediate post-war era in which civil aviation would play an essential role in forging a new global economic and trade order, including between nations formerly at war. It was the essence of that transition from devastating war to a peaceful and prosperous future that weapons of war (aircraft) could be repurposed for the movement of people and goods around the world based on an orderly, globally accepted system of rules, reciprocal recognition, and mutual accommodations among nations. As the Convention's preamble states: "the future development of international civil aviation can greatly help to create and preserve friendship and understanding among nations…to avoid friction and to promote the cooperation between nations…upon which the peace of the world depends."[96]

The Chicago Convention covers a wide range of topics, including the sovereignty of States over their own airspace and the rights of aircraft of one State to overfly the territory of other States, to make technical stops in other States, and to take on and discharge passengers and cargo on a charter basis at airports in other States. The Convention also addresses regulation of aircraft by nationality (the State in which it is registered), air navigation, licensing and certification of aircraft and crew, the development of safety standards and practices, and the settlement of disputes between States.

ICAO

The Chicago Convention established ICAO as an international governing body for civil aviation. ICAO's main functions include (1) developing and revising matter-specific Annexes to the Convention that establish Standards and Recommended Practices (SARPs) for aviation safety and security, (2) addressing issues of access to airspace and airports in other countries, (3) serving as a clearinghouse for cooperation and discussion on civil aviation issues, and (4) providing a forum and procedures for resolution of disputes between States.

Evolution

Over time, ICAO has sought to implement the Chicago Convention's commitment to create a unified post-war era civil aviation sector, with a primary focus on aviation safety and security. As described below, ICAO has had important successes but has also struggled with significant challenges.

ICAO's main achievements

Over the past nearly 80 years, ICAO has proven its durability. Its greatest successes have been in aviation safety. ICAO's status as a UN body underscores its authority to bring Member States together to address often-complex safety problems. ICAO has developed a modus operandi whereby Member States can participate at a high level in initially establishing policy objectives and ultimately approving specific measures for global implementation, while leaving the technical "sausage making" of SARP developments to industry experts who work on the details in a less politicized (but never entirely apolitical) environment. ICAO's workings are relatively transparent and based on cooperation among Member States, all of whom have a vested interest in global aviation safety and the relatively free movement of aircraft. The following are examples of SARPs that Member States have implemented:

- The establishment of standards for an airborne traffic alert and collision avoidance system that interrogates air traffic control transponders in nearby aircraft and uses computer processing to identify and display potential and predicted collision threats (i.e., the automated system that alerts a pilot in flight to "pull up" in response to a risk of collision).[97]

- The development of standards for Flight Data Recorders (FDRs), which provide critical information for investigators in understanding why an aircraft crash may have occurred.[98] Member States, which often cooperate on accident investigations, have a strong common interest in the gathering and preservation of FDR data in the event of an accident, so the establishment of uniform FDR standards continues to be of great importance for ICAO.

- The creation of principles and instructions governing the international transport of dangerous goods by air, such as the now ubiquitous transport of highly flammable lithium batteries onboard civil aircraft.[99]

- The creation of the Safety Management System (SMS) State Safety Program (SSP),[100] which set forth comprehensive, systematic, and cohesive approaches to managing safety (i.e., structures, accountabilities, policies, and procedures). The FAA and other Member State regulators now require SMS compliance for all large commercial air carriers.

- The development of aircraft noise standards, which provide maximums for the noise levels that civil aviation aircraft may emit. These standards

have been adopted by the FAA for the new type certification of jet and turboprop aircraft.[101]

ICAO's main challenges

The challenges ICAO faces include the inherently political nature of governance, deliberation, and compromise among 193 nations. Because ICAO lacks enforcement authority, it relies on Member States to comply with the technical guidelines it produces. In practice, enforcement occurs bilaterally and multilaterally between and among Member States. ICAO's processes can be hamstrung by bureaucracy as well as intergovernmental politics. This impedes ICAO's ability to respond more nimbly and effectively to urgent aviation safety problems. For example, it falls to individual Member States to "ground" aircraft in response to safety problems (e.g., the Boeing 737 MAX)[102] or impose specific retaliatory or restrictive measures on a Member State (e.g., the response to Russia's invasion of Ukraine).[103]

ICAO also has struggled (but arguably has achieved some success based on international compromise) to develop a global approach to commercial aircraft emissions, which account for about 2.5% of global carbon emissions. After the EU grew impatient with the pace of progress to address the issue at ICAO, it developed its own initiative, an Emissions Trading Scheme (ETS), that would apply to aircraft of non-EU Member States.[104] ICAO's compromise, the so-called Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA) provides for a multi-year, phased process for Member States to meet certain limits on aircraft carbon dioxide emissions, culminating in net-zero emissions by 2050.[105]

CORSIA remains controversial, however, with the EU threatening to reinstate the ETS if CORSIA is not implemented on schedule.[106] China and Russia, by contrast, have refused to commit to participate in Phase One of CORSIA (which will run through 2026 and for which participation is voluntary), while maintaining that they will participate in Phase Two (which will begin in 2027 and for which participation will be mandatory).[107] China and Russia argue that a requirement to meet certain targets within CORSIA's timeframes would unfairly penalize developing countries.[108] China's refusal to fully participate in CORSIA could make it more difficult to ensure the participation of other countries.

While ICAO has ultimately achieved an effective role in safety regulation, it lacks a similar role in the areas of economic/trade and security relations among

nations relating to air transportation. Nations generally negotiate bilaterally to exchange scheduled air service "traffic rights," which has produced a system that lacks uniformity and arguably is excessively protectionist (e.g., the airline industry remains subject to varying restrictions on foreign or cross-border ownership, which do not apply to most other global industries). Nations have also adopted a more unilateral approach to aviation security, with the events of September 11, 2001, having accelerated that trend.

For example, the United States has established its own specific requirements for passenger and cargo security screening. If an airline of a foreign country that is also a Member State wishes to fly passengers to the United States, it must gather and transmit specific passenger data to US authorities in advance of the flight and submit the aircraft and its passengers to US screening requirements. If a foreign airline or its government refuses to comply, the United States may refuse entry to that airline—regardless of the Convention's provisions on providing access to airspace and airports. Other Member States have established their own security screening and entry requirements.

Governance

ICAO's governance structure

ICAO has three main bodies that serve to carry out its mission and purpose: the Assembly, the Council, and the Secretariat.

- The Assembly is ICAO's supreme body and is composed of delegations from ICAO's 193 Member States. The Assembly meets every three years to set ICAO's agenda, vote on major policy initiatives, and elect Member State representatives to the Council. Industry and civil society groups, along with various regional and international organizations, also participate in these events in their capacity as "Invited Organizations."

- The Council is ICAO's governing body, comprising of representatives from 36 Member States appointed by the Assembly to serve three-year terms. After the Assembly approves a policy initiative, the Council convenes expert panels and working groups to develop a SARP. These industry experts may be recommended by Member States but do not represent the interests of any particular State; rather, they provide objective technical expertise and recommendations on how best to address a particular safety issue. Any new

SARP recommended by an expert panel is subject to review by the Secretariat (see below) and approval by the Council and ultimately the Member States through the Assembly. In recent years, the Council also has developed aircraft CO2 emissions reductions measures, at the request of the Assembly.

- The Secretariat is ICAO's professionally staffed executive body. It is led ICAO's Secretary General and is responsible for managing ICAO's day-to-day operations.

SARPs

SARPs are the primary tool for implementation of ICAO-approved safety standards and practices. "Standards" are presumptively mandatory: specifications "the uniform application of which is recognized as necessary for the safety or regularity of international air navigation and to which…States will conform in accordance with the Convention."[109] "Recommended practices," meanwhile, are hortatory: specifications "the uniform application of which is recognized as desirable in the interest of safety, regularity or efficiency of international air navigation, and to which…States should endeavor to conform in accordance with the Convention."[110]

SARPs may address the full range of subjects covered by the ICAO Annexes, including pilot and crew licensing, rules of the air, meteorological services, air navigation and air traffic control services, safety management, aircraft operations, aircraft airworthiness, aircraft nationality and registration, search and rescue, accident and incident investigation, airport regulation, the transport of dangerous goods by air, and environmental protection and security issues.

The ICAO Council, which meets three times annually, may propose a safety issue for review. (Such a proposal may also originate in the ICAO Assembly, which may direct the proposal to the Council.) The Council then refers a proposal to ICAO's Air Navigation Commission (ANC).

The ANC is comprised of 19 members who are nominated by Member States and appointed by the Council. The ANC has 17 technical panels with specific subject-matter expertise (e.g., safety management, remotely piloted aircraft systems, dangerous goods). The relevant ANC technical panel will then conduct research as a basis for potentially drafting a SARP for the ANC's review. If the ANC decides that the SARP is warranted, the ANC will finalize the SARP, consulting

informally with the Secretariat (while the Secretariat's approval of a SARP is not required, the Secretariat provides technical, legal, and administrative support). The ANC then submits the proposed SARP to the Council where adoption requires the approval of two-thirds of the Council's members. Thereafter, the SARP is distributed to the Member States, which have three months in which to approve or disapprove the SARP.

Unless a majority of Member States register their disapproval, the SARP becomes effective four months after its adoption by the Council. Member States may lodge "differences" with ICAO (i.e., the intention of a Member State to deviate from some aspect of the SARP), however, practically speaking a Member State that has notified a difference is motivated to eventually harmonize its national regulations, as one State's failure to conform to a particular standard may form a basis for other States to eventually withhold approvals for the non-conforming State's aircraft operators. After ICAO adopts a SARP, Member States are charged with implementing it into their national laws and regulations. This process varies from State to State. In the United States, the FAA (or another federal agency, as may be applicable) generally incorporates SARPs directly into its regulations. For example, after ICAO adopted a SARP regarding aircraft engine emissions, the Environmental Protection Agency (EPA), which regulates engine emissions, conducted a rulemaking to incorporate the SARP into its regulations. US legal and policy requirements pertaining to agency rulemaking (e.g., public notice and comment requirements) may delay full US implementation of a SARP. Member States also pursue uniformity of SARP adoption and implementation via bilateral and multilateral (e.g., regional) aviation safety agreements.

Broader global governance landscape

Bilateral aviation safety agreements

The United States and other Member States have entered into bilateral aviation safety agreements (BASAs) in an effort to achieve: 1) broader compliance with ICAO Annexes and SARPs; and 2) as a related matter, a greater degree of consistency between the safety regulations of Member States. BASAs provide for bilateral cooperation in a wide variety of safety areas, including aircraft and crew licensing, air navigation, aircraft maintenance, and flight operations. BASAs often reference and incorporate SARPs or, more generally, adherence to ICAO standards. The United States and other Member States use BASAs as a way to

harmonize their respective safety regulatory frameworks. In some cases, such as between the United States and the European Union, each Party may defer to the other's licensing, compliance, and other safety determinations. As Article 5 of the US-EU BASA states: "[T]he Parties agree that each Party's civil aviation standards, rules, practices and procedures are sufficiently compatible to permit reciprocal acceptance of approvals and findings of compliance…"[111]

Dispute resolution

Under chapter XVIII of the Chicago Convention, the ICAO Council provides a forum for the resolution of disputes between Member States relating to the interpretation or application of the Convention and its Annexes. In practice, however, such disputes are rarely brought to ICAO and are even more rarely adjudicated. This is because bilateral air transport agreements between Member States generally include rights and procedures both informal (e.g., intergovernmental consultations) and formal (e.g., arbitration) that offer a more direct and efficient path to dispute resolution.

Under ICAO dispute resolution procedures, Member States must first attempt to resolve a dispute by direct negotiation. Only after failed negotiations may a Member State seek resolution by a decision of the ICAO Council. A Member State may appeal the Council's decision to an ad hoc arbitral tribunal or the Permanent Court of International Justice. The ICAO dispute resolution process is protracted and slow moving. In most cases, Member States resolve a dispute before the Council renders a decision, but in some cases a Member State may submit a dispute to ICAO in an effort to apply additional pressure on another Member State to resolve the matter.

ICAO does not have direct authority to impose sanctions regarding the specific subject matter of a dispute, but individual Member States may use a Council decision as a basis for refusing access to its airspace or territory. The ICAO Assembly may suspend the voting rights of a Member State in the Assembly following a Council decision that the Member State is in "default" of its obligations under the Convention.

Compliance and enforcement

ICAO does not directly enforce SARPs; rather, it falls to Member States, individually and via bilateral and multilateral agreements, to ensure compliance.

ICAO, however, plays a role in "assisting" Member States to comply with ICAO's Annexes and SARPs, including by conducting safety audits of Member States. ICAO's auditors examine Member States' legislation and regulations for compliance with ICAO Annexes and SARPs. ICAO's audit reports, which are published on ICAO's website, identify any significant safety concerns. ICAO does not conduct audits of airlines or airports; such regulation falls to the civil aviation authorities of individual Member States.

Although ICAO does not have authority to enforce compliance with its Annexes and SARPs, Member States may use information and findings contained in ICAO audit reports to improve their safety oversight regimes. Some Member States also audit other states' compliance with ICAO standards and impose restrictions on access to national airports and air service markets based on a finding of deficient compliance. The United States and the EU have adopted different approaches to auditing Member States' compliance with ICAO standards. The FAA has established an International Aviation Safety Assessment (IASA) program under which it audits and then assigns ratings to other Member States, either a Category 1 rating (complies with ICAO standards) or Category 2 rating (non-compliant). The EU, by contrast, asks countries to audit themselves to confirm their compliance with ICAO standards. The EU maintains a blacklist of airlines determined to have serious safety deficiencies, prohibiting those airlines from operating to or within the EU.

The FAA's IASA program's audits and country ratings have a significant impact on international commercial air transportation because the United States is the world's largest air service market. For example, in May 2021, the FAA downgraded Mexico from a Category 1 to Category 2 rating following an FAA audit finding that Mexico was not in compliance with ICAO standards. Consequently, the FAA prohibited Mexican airlines from introducing new services to the United States or engaging in codesharing with US airlines, where a US airline would sell tickets for travel on a Mexican airline under the US airline's two-letter code. The FAA allowed Mexican airlines to continue operating services to/from the United States that were already in place at the time of the downgrade. In September 2023, the FAA restored Mexico to Category 1 status. In doing so, the FAA noted that "[w]ith a return to Category 1 status, [Mexican airlines] can add new service and routes to the US, and US airlines can resume marketing and selling tickets with their names and designator codes on Mexican-operated flights."[112]

The FAA, in announcing the restoration of Mexico's Category 1 rating, emphasized how the FAA had made its "expertise and resources" available to

provide "technical assistance" to enable Mexico's civil aviation authority to achieve compliance with ICAO standards.

Conclusion

To paraphrase Winston Churchill, ICAO, like democracy, is the worst possible governance system—except for all of the alternatives. Although imperfect and limited, particularly in non-safety areas, the ICAO regulatory scheme enabled the post-World War II development of a global air transport industry in which weapons of war (aircraft) were converted into vehicles for the safe global movement of people and goods, for the greater economic and social benefit of the world.

In some respects, ICAO's greatest success is its endurance. It has survived for nearly 80 years and there is no discussion about replacing or abandoning it. ICAO will likely endure and continue to provide leadership in the essential area of aviation safety for the foreseeable future. In other areas, however, nations are likely to forge ahead based on unilateral action (e.g., security) or initiatives that are the product of regional coordination or understandings between nations (e.g., the exchange of air traffic rights and the related issue of rules governing the ownership and control of airlines).

The environment may prove to be a bellwether of ICAO's future. While ICAO has touted CORSIA as "the first time that a single industry sector has agreed to a global market-based measure in the climate change field," it represents an uneasy compromise between nations that want to move more quickly or slowly to address aircraft emissions. If that compromise does not hold on what has become one of the most challenging points of controversy in international aviation, Member States may revert to unilateral approaches, which in turn could undermine ICAO's authority and effectiveness as an aviation safety regulator.[113]

3.2 The European Organization for Nuclear Research (CERN)

Authored by Professor Sir Christopher Llewellyn Smith

Purpose

An organization with 23 Member States (22 European and Israel), CERN seeks to advance the boundaries of human knowledge through research in particle physics.[114] Originally an acronym for Conseil Européen pour la Recherche Nucléaire, CERN now styles itself the European Laboratory for Particle Physics. CERN constructs and operates facilities that are used by over 13,000 physicists from around the world (the "users") and employs around 3,390 fellows and permanent staff. Many of the components of CERN's large particle detectors are largely built in the users' home institutions and then transported to CERN.

CERN hosts the Large Hadron Collider, the largest and highest-energy particle collider in the world. The laboratory has made major contributions to current understanding of the structure of matter and invented, developed and pioneered the use of a wide range of technologies, the best-known examples being the discovery of the Higgs boson and the invention of the World Wide Web.

CERN was conceived in the late 1940s with the dual aims of enabling the construction of facilities beyond the means of individual countries—thereby allowing European physicists to compete with their peers in the USA, where large accelerators were being built—and fostering cooperation between peoples recently in conflict.

From the outset, CERN intended its findings to be widely accessible. CERN's equivalent of a constitution, its Convention, stipulates that "the Organisation shall have no concern with work for military requirements and that "the results of its experimental and theoretical work shall be published or otherwise made generally available". CERN shares its technology and knowledge with companies and research institutes, and its experts frequently consult with businesses. CERN encourages the creation of new companies based on its technologies and grants licenses to commercial and academic partners for the use of its technologies.

Patents are only filed if doing so makes technologies more attractive to companies interested in using them.

History

At a meeting of the United Nations Educational, Scientific and Cultural Organization (UNESCO) in Paris in 1951, 12 European governments adopted a resolution establishing CERN (CERN is not part of the UN system, and although UNESCO has been an Observer since the beginning, it did not send representatives to meetings of the CERN Council for many years). Two months later, an agreement created CERN's Provisional Council, which drafted the Convention that governs CERN. The Convention was signed by the original 12 Member States in June 1953, and CERN formally came into existence on September 29, 1954 when it had been ratified by all twelve Members.

In 1952, the Swiss, Dutch, French, and German governments submitted proposals to host the CERN laboratory. Geneva was ultimately chosen due to its central location and Switzerland's neutrality in World War II. While technical factors (such as the availability of large amounts of electrical power) can be helpful in making shortlists of potential sites for international organizations, the experience of CERN and other similar organizations indicates that political and economic factors tend to dominate.[115] Factors to be considered, apart from money, when selecting a site include: logistical ease of access, openness to visitors, accommodation, and schooling.[116]

Throughout CERN's history, collaboration has created connections that cross political and cultural divides and foster better international understanding. CERN was the first intergovernmental organization that Germany joined after World War II. During the Cold War, CERN maintained links with scientists behind the Iron Curtain. In the 1980s, CERN became one of the first European scientific organizations to welcome significant numbers of Chinese scientists.

Evolution

CERN's facilities have grown enormously over the years, and today it is the world's pre-eminent laboratory for particle physics. The number of users has also grown, although they currently look set to decline following the CERN Council's announcement that cooperation with Russia and Belorussia will come to an end

when the current agreements expire in 2024. While CERN has grown spectacularly, the individual Members' contributions to the budget have remained roughly constant or even declined in real terms.

Since its inception, CERN has also grown from 12 to 23 Member States, mainly because of the accession of formerly communist countries. In CERN's early years, Observers (which included both organizations, such as UNESCO and the EU, and non-member countries) received invitations to attend public sessions of the Council. While not entitled to speak, Observers may be invited by the President to do so.

During the Large Hadron Collider (LHC) construction era, non-European countries contributing 15 million Swiss Francs or more to its construction were granted Observer status, which came with the right to contribute to the LHC decision-making process. This "Observership with special rights" was granted to four states (Israel, Japan, Russia, and the United States). In 2010, this status was replaced by a new Associate status, and it was decided that the status of Observer should be granted only to organizations.

Today there are nine Associate Member States, including three (Cyprus, Estonia, and Slovenia) in a pre-stage to membership. Their annual contributions are set at a level that is high enough to have a tangible impact on the CERN budget without discouraging applications. Associate Member States are granted the right to attend the Council's open and restricted (but not closed) sessions and can send representatives to finance committee meetings. They cannot vote in the Council and its committees but can ask for the floor and make statements without having been invited to do so.

Governance

CERN is an intergovernmental organization, established by a treaty, that possesses its own international legal personality. Changing the Convention, which provides the framework for the organization's governance, is difficult. It requires unanimity and ratification by all Members, which typically involves approval by their national legislative bodies. This has proved to be a source of stability. CERN has only revised its convention once, in 1971, when it established a substantial presence in France in addition to Switzerland.

The CERN Convention, which has served CERN well for nearly 60 years despite significant changes in its size and nature, reflects the long-term vision of CERN's founders and grants the Council powers that have provided important flexibility. It has, for example, allowed Israel to become a Member State, despite the word "European" appearing in CERN's official title.

The Convention's flexibility is one of the pillars on which CERN's success rests. The other is the trust that Member States have in the laboratory's management and technical judgements. There has only been one major review of CERN's management, which was carried out in the 1980s as a condition for the UK's continued membership after it had considered withdrawing.[117] In contrast, historians of the US Superconducting Super Collider attribute its demise partly to almost continuous management and technical reviews by the Department of Energy.

Alternatives to treaties

Signing onto international treaties generally requires legislative or parliamentary approval. In the case of the US, joining international organizations or collaborations not established by treaty is normally "subject to the annual availability of funding", which creates unease among other parties that have made long-term commitments. There are several examples of international scientific organizations with alternative structures.

- The Institut Laue–Langevin (ILL) in Grenoble, which houses a high flux nuclear reactor that is used to study materials on short-distance scales, is a private company under French law that is jointly owned and governed by French, German, and UK scientific organizations. They work closely with the ILL's 11 European "Scientific Member countries", who together contribute some 20% of the annual budget.

- Similarly, a nonprofit limited liability company owned by participating countries is responsible for constructing and operating the European Xray Free Electron Laser (XFEL), based at the DESY laboratory in Hamburg. Likewise, the Facility for Antiproton and Ion Research (FAIR), which is an international center and one of the world's largest research projects, is being built by a private company at GSI. Both DESY and GSI are large, established laboratories onto which XFEL and FAIR are being grafted.

- The Joint European Torus (JET) at Culham in the UK provides another model. About 350 scientists from EU countries and other countries from around the globe participate in JET experiments each year under the scientific direction of a leader appointed by Eurofusion. The Culham Centre for Fusion Energy (CCFE) is responsible for maintaining and upgrading JET, under a contract between the European Commission and the UK Atomic Energy Authority (CCFE's operator). This funds around 400 engineers and technical staff who are responsible for operating and maintaining JET.

Structure and leadership

CERN's governing body, the Council, is composed of two delegates from each Member State. Typically, one of the delegates is a government representative (often from a Ministry of Science, or in some cases, the country's ambassador to the UN organizations in Geneva) and the other is a scientist. This combination of political and technical representation has served CERN well. The Council elects a President and two vice Presidents from among the Delegates and appoints the Director-General, who is the chief executive officer of the Organization and its legal representative. The Convention stipulates that in the discharge of his or her duties, the Director General "shall not seek or receive instructions from any government or from any authority external to the Organization".

The Convention established a Scientific Policy Committee (SPC) and a Finance Committee (FC). The SPC's mandate includes setting research priorities, measuring CERN's achievements against annual goals, and overseeing senior staff appointments. Its members include individuals of the highest standing in the scientific community, who are appointed by the Council, and the Chairs of various advisory committees. All act as individuals, not as national representatives (the members include nationals of non-member states) or as representatives of the bodies they chair. The FC provides the Council with advice on financial matters, approves large-scale contracts and staff regulations, and recommends staff rules to the Council.

The "President's group," which includes the Director General, the two Vice Presidents, and the chairs of the finance and scientific policy committee, helps the President prepare for Council sessions. There is a tradition that during its meetings, which normally take place over dinner between meetings of the FC and of the Council, the Delegates take off their hats as national representatives and discuss how to conduct the business in what they consider to be the best interest of CERN.

Voting

While unanimity is required for changes to the Convention, admission of new members, and approval of major projects, almost all other matters are in principle decided by a two-thirds majority (some international organizations require unanimity for most decisions, which is known to have led to difficulties in some cases). However, CERN has a long tradition of reaching consensus on difficult issues through diplomatic means, such as informal negotiations between delegates, rather than formal voting, and has effectively abandoned the two-thirds majority rule for major financial issues.

In its first decades, when CERN had 12 Member States, there was a tacit understanding that countries that made relatively small financial contributions would not outvote a majority of members that made major financial contributions on important financial issues. In 1991, when there were 16 Member States, it was decided that the FC's recommendations to the Council should be backed by 55% of the annual financial contributions of the Member States, in addition to the majority required by the Convention. This number was later increased to 70%.

CERN's plans include the possible construction of a 90-km circumference Future Circular Collider by a large global collaboration of partners. A specially constituted Council Working Group on the Governance of CERN is currently considering how such a project might be governed.

Funding

While non-member States have contributed to the construction of the LHC in-kind, Members' regular contributions to the budget of CERN are all in cash. Cash contributions with open bidding for contracts leads to lower costs. At the International Thermonuclear Experimental Reactor (ITER), where the major contributions are in-kind, construction of some large components has been split between suppliers in different member states. This has produced technical issues and raised costs, as each supplier incurs their own set-up costs.

The CERN Members' contributions are calculated as a percentage of their average net national incomes for the preceding three years. Until the late 1980s, "average" was interpreted as a simple average, and Members' payments reflected their past—rather than their current—economic strength. Since then, CERN has used weighted averages that account for trends in relative economic strengths and changes in exchange rates.

The original Convention set a maximum percentage for the contribution of any Member. This was removed when the Convention was revised in 1971, but the Council subsequently set a maximum. This protects the biggest contributors from feeling that they are carrying the main burden without receiving more influence. The Council can also take into account a Member State's situation and temporarily reduce its contribution, as it is currently doing with Ukraine, which is an Associate Member.

Countries that host international organizations benefit from staff salaries being spent locally, as well as the placement of most small and service contacts. Consequently, the hosts of some organizations are required to pay a "host state premium". France and Switzerland, CERN's two host nations, have made additional voluntary contributions, some of which were in-kind.

Procurement

CERN calculates a return coefficient, which is the ratio between a Member state's percentage share of the value of all contracts and its percentage contribution to the CERN budget. Members are said to be "poorly balanced" if their return coefficient is less than 1.0, and "well balanced" if it is greater than or equal to 1.0. When awarding new contracts, consideration is given to whether the lowest bidders are well or poorly balanced. If the lowest bid is from a manufacturer in a well-balanced country, then the two lowest bidders from poorly balanced countries are offered the opportunity to adjust their bids to match the lowest bid, as long as their bids were within 20% of that bid.

Conclusion

In its mission of advancing the boundaries of human knowledge through research in particle physics, CERN has been a success. Analysis of the way that CERN and other international scientific organizations referenced above work leads to a list of issues that will have to be addressed in establishing new international organizations, including their legal status, voting procedures, the basis for calculating contributions, the constitution of advisory bodes, and site selection. How best to deal with these issues will depend on an organization's mandate. Issues that deserve particular attention in establishing an organization charged with governing AI (which, unlike CERN, will presumably not be a user organization, and will not require infrastructure that will take decades to construct) include:

Intellectual property, openness and independence. CERN's core tenet of separation from military endeavors and the accessibility of its scientific research has been central to its mission, as has the stipulation that the Director General of CERN's laboratory operates independently of any government or outside institution. Questions of independence and accessibility will be similarly critical with regard to an AI governance organization.

Whether to graft a new organization onto an existing body. The core of an AI organization will likely be its staff, supported by computing power which could presumably be acquired relatively quickly compared to constructing a new fusion device or accelerator. Grafting such an organization onto an existing body (as XFEL, FAIR, and JET have done) would allow it to rely largely on pre-existing administrative support and services and get off to a rapid start.

The possible involvement of private companies. Creating an international organization to which private companies belong alongside countries would raise novel governance issues. If private companies are formally involved in an AI governance organization, these issues might be finessed by making these companies Observers or giving them some sort of associate status.

The possible involvement of a politically neutral "parent body" such as UNESCO, to which all potential members already belong, lowering political barriers to joining. An example is provided by SESAME (Synchrotron-light for Experimental Science and Applications in the Middle East), whose Members include Iran, Israel and Palestine, which (like CERN) was founded after UNESCO summoned a meeting of potentially interested parties (in contrast to CERN, UNESCO continues to play a role in SESAME). UNESCO's involvement made it easier for some countries to participate than it might otherwise have been.

Evaluation of the benefits of cash and in-kind contributions from member states. In the case of an AI institute, the major purchases will presumably be of computing resources. In this case, the organization could be funded by cash contributions, which would purchase equipment or services on the basis of open tender. While using fewer vendors or even a single vendor would improve technological compatibility, this would raise the issue of industrial returns to members.

The development of a new international governance organization offers the opportunity to learn from the challenges faced and obstacles overcome by CERN. These lessons have the potential to help accelerate science and advance human potential in the field of AI and beyond.

3.3 The International Atomic Energy Agency (IAEA)

Authored by Dr. Trevor Findlay

Purpose

The International Atomic Energy Agency (IAEA) is a multilateral, intergovernmental organization that pursues a variety of interrelated governance[118] missions, including nuclear safeguards, nuclear safety and security, and technical assistance with nuclear technology. Established in 1957 in Vienna to promote and govern the peaceful uses of nuclear energy worldwide, the IAEA is best known for the nuclear safeguards system later put in place and its unparalleled monitoring, verification and compliance capacities.

The IAEA's safeguards system represents the most radical impingement on national sovereignty yet devised: safeguards are legally binding for most states, they encompass extensive monitoring and verification measures (including notably intrusive, mandatory on-site inspections), and the Agency has direct access to the United Nations Security Council to request enforcement measures. IAEA also defines safety and security standards for handling nuclear technology and helps developing countries identify energy needs and use nuclear technology.

The success that the IAEA has helped achieve in avoiding nuclear catastrophe on a global scale offers lessons that may be applicable to the creation of a new governance organization. For example, the IAEA encourages states to accept impingements on their sovereignty in return for an orderly regime that benefits all states. It also offers assistance to states regarding the peaceful uses of nuclear technology to enhance this "bargain."

When it comes to constructing a new international regime, the bargain struck between developed and developing countries can mean the difference between success and failure. Such a bargain may involve enhanced regulation, monitoring, verification, and compliance mechanisms in exchange for development assistance and technical cooperation.

History

The objective of the IAEA, as set by its statute, is as follows:

> ...to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world. It shall ensure, as far as it is able, that assistance provided by it or at its request or under its supervision or control is not used in such a way as to further any military purpose.

The establishment of the IAEA stemmed from US President Dwight D. Eisenhower's "Atoms for Peace" speech at the UN General Assembly on December 8, 1953. Eisenhower suggested creating an agency that would receive nuclear material from "advanced" nuclear nations and provide this material to member states for peaceful use in medicine, agriculture, science, and power generation. The hope was that this clearinghouse arrangement would not only decrease the stocks of nuclear material available for nuclear weapons, but also head off aspirations by additional states to acquire such weapons.

Following secret talks between the US and the Soviet Union, a select group of states convened in Washington, DC to negotiate a draft statute. This statute was subsequently amended and adopted by the UN General Assembly in 1956, and the Agency was established the following year.

In many respects, the IAEA was an American project—initiated, developed, and funded generously by successive US administrations until it took on a life of its own. When creating a new governance organization, it is often helpful if a policy leader (or a coalition of them) emerges quickly to drive the process, as a negotiating free-for-all will likely not produce the necessary coherence and effectiveness.

Evolution

The original concept for the IAEA as a nuclear material clearinghouse never eventuated, partly because more states began to start their own nuclear programs.

The US also began to directly supply other countries with nuclear assistance, under bilateral US safeguards agreements to prevent misuse for weapons purposes. The Soviets soon followed with their own program.

The IAEA instead became the "nuclear watchdog", establishing a nuclear safeguards system to deter states without nuclear weapons from manufacturing them. In the 70 years since its inception, the IAEA has also adopted new governance roles in nuclear safety (preventing nuclear accidents) and nuclear security (preventing nuclear terrorism). In addition, the IAEA provides technical assistance to member states in a manner that vastly exceeds what was envisaged in the statute.

The nuclear safeguards system

The IAEA's regime for detecting the diversion of peaceful nuclear materials to weapons purposes is known, confusingly to outsiders, as "safeguards". The safeguards system involves states declaring to the Agency the types, amounts, and locations of nuclear materials in their possession. The most sensitive materials are enriched uranium and plutonium, both of which may be used for nuclear weapons, and both of which also feature in a sophisticated nuclear fuel cycle designed for peaceful purposes.

The Agency applies several layers of safeguards measures to ensure that state declarations are correct, including:

- Nuclear material accountancy;
- On-site inspections (the Agency employs roughly 200 inspectors to carry out on-site activities, as well as a cadre of information analysts and technical support staff);
- Seals to ensure that material is not tampered with between inspections;
- Sample analysis;
- Remote video monitoring;
- Satellite imagery;
- Open-source information analysis; and

- In extreme circumstances, the analysis of intelligence information provided by member states.

In theory, at least, the consequences for a state caught in non-compliance are serious. Once the IAEA director general reports a non-compliant state to the UN Security Council, the Council is empowered to punish such violators with sanctions, including economic sanctions, and ultimately, the use of military force.

The system has been subject to almost continuous technical improvement since being established in the late 1950s. Originally, safeguards were purely voluntary, imposed as states offered nuclear materials to others and wished to have reassurance that such material would not be misused. A major shift occurred in 1970 with the entry into force of the 1968 Nuclear Non-Proliferation Treaty (NPT). The NPT made IAEA safeguards mandatory and legally binding for states without nuclear weapons, but not for the five official nuclear weapon states—China, France, the Soviet Union/Russia, the United Kingdom and the US.

Non-nuclear weapon states were obliged to sign bilateral agreements with the IAEA establishing the scope and nature of their safeguards obligations, which varied depending on national circumstances. The NPT vastly increased the importance and technical capacities of the IAEA and its safeguards system.

After the discovery in 1991 of an illicit Iraqi nuclear weapons program, the IAEA further strengthened and modernized nuclear safeguards by negotiating an Additional Protocol (AP) for bilateral safeguards agreements between states and the IAEA. The adoption of an AP by states is voluntary, although a substantial majority of states have chosen to adopt one.

The safeguards regime, both by design and by accidents of history, creates different obligations for different states, which has led to charges of inequity and discrimination:

- As the IAEA was established two decades before the NPT for a different purpose, not all IAEA member states (notably India, Pakistan, and Israel) are party to the NPT, yet these states may still be elected to the board of governors and sit in judgement on other states violating their NPT safeguards obligations;
- Not all IAEA member states or parties to the NPT are required to have safeguards agreements (the nuclear weapon states are only encouraged to adopt "voluntary" agreements); and

- Not all states with safeguards agreements have concluded a voluntary Additional Protocol, the highest level and most intrusive form of safeguards (notably Argentina, Brazil, Egypt, Iran, Saudi Arabia, and Syria).

Ensuring that the establishment of an international agency flows directly from its foundational treaty is one way to avoid such complexities. This is the model followed by more recent examples, such as the Comprehensive Nuclear Test Ban Treaty Organization, established pursuant to the 1996 Comprehensive Nuclear Test Ban Treaty, and the Organization for the Prohibition of Chemical Weapons, established pursuant to the 1993 Chemical Weapons Convention.

The nuclear safety and security regimes

While the nuclear safety and security regimes are also based on legally binding treaties, they are not subject to the legally binding reporting, monitoring, verification, and compliance processes of the safeguards regime. Often referred to as "incentive regimes", the treaties only commit states to making their best efforts to achieve safety and security. The measures applied include voluntary reporting, recommended standards and practices, assessment missions, periodic review conferences, and technical assistance.

The rationale behind the safety and security regimes is that states themselves should have primary responsibility for the safety and security of their nuclear enterprises, and the IAEA should only advise and assist them in carrying out such tasks.

Like safeguards, these regimes have become more extensive and sophisticated in response to clarifying events, including the accidents at Chernobyl (1996) and Fukushima (2011). None of the innovations that followed these crises included intrusive measures, such as on-site inspections.

The IAEA has found that the development of agreed standards and codes of conduct, even if not mandatory, can have a normative effect. The downside of these measures is that agreement on standards and recommendations tends to devolve to the lowest common denominator. Additionally, the IAEA uses visiting missions, comprising both IAEA and national representatives, to assess implementation and make recommendations. Over the years, this has led to improvements in state performance.

Governance

Any UN member state may join the IAEA. As of September 2023, the Agency had 178 members out of 193 UN member states, making it close to universal (mostly small island states are unrepresented). All states possessing nuclear weapons or with significant peaceful nuclear activities are members, with the stark exception of North Korea, which withdrew in 1994—the only state ever to have done so.

Achievements and challenges

The IAEA confronts the classic dilemma of all international organizations— it is both empowered and hindered by its member states. The Agency is crucially dependent on states to carry out its mandate on their behalf. This means that the director general and secretariat can only act with the approval and support of member states, especially the most powerful. The United States, for instance, provides up to 25% of the IAEA's regular budget, in addition to generous voluntary contributions and technical assistance.

China, Russia, the European Union, and developing countries collectively have also become key players. Such power dynamics are especially prominent in determining Agency action against states that have violated their safeguards obligations. They also arise when the secretariat attempts to further strengthen safeguards.

On the other hand, the IAEA, like other international organizations, has carved out a certain autonomy in the nuclear field. The increasing complexity of the nuclear enterprise, the number of industrial players that have emerged, and the expansion of IAEA membership means that only a handful of states can keep track of all the IAEA's activities and acquire the same familiarity with global nuclear governance as the Agency itself. In carrying out its mission, the IAEA has also attempted to portray itself—not always successfully, given the political issues at stake—as a science and technology based institution that is impartial, autonomous, and non discriminatory in its dealings with its member states.

The IAEA levers its accumulated experience and expertise to establish and reinforce good behavior in all areas of its mandate. It can produce compromises among its member states by trading off their competing interests against each other. A recent example is the "7 Pillars of Nuclear Safety and Security" that the

current Director General, Rafael Grossi, issued immediately following Russia's invasion of the Chernobyl and Zaporizhzhia nuclear power plant sites.

The role of industry

From the beginning, the IAEA has kept the industry it was supposed to be governing at arm's length, a flaw that has long been apparent but only recently addressed. This is due in part to the fact that the IAEA's establishment was driven by the concerns of national leaders about the dangers of nuclear weapons proliferation.

The impetus did not come from the nuclear industry, which barely existed in the 1950s and was almost exclusively operated by governments.

The IAEA has historically handled nuclear governance matters via either member states' permanent diplomatic representatives in Vienna or foreign offices in member states' capitals. While some delegations, notably those of China, Russia, and the US, include nuclear experts, these are mostly from national nuclear bureaucracies, such as the US Department of Energy; government-run nuclear laboratories, such as Sandia in the US; or national regulators, such as the US Nuclear Regulatory Commission. Nuclear industry has not been invited to join national delegations to IAEA conferences, the theory being that companies can interact directly with their national governments to protect their interests.

For their part, private companies in the nuclear field have also tended to keep their distance from the international regime. They almost invariably regard the IAEA and governments as seeking to intrude on their commercial operations and see instruments such as the NPT as "political" documents of no concern to them. From the outset, the privately owned uranium mining industry pressured governments whose territory contained large uranium deposits (such as Australia, Canada, and Belgium) to exempt natural uranium from IAEA safeguards.

Today, industry is more involved in the nuclear security issue, presumably due to the commercial implications of a catastrophic nuclear terrorism incident. This has led, for instance, to industry-organized summits that coincided with the state-led Nuclear Security Summits held at US initiative from 2010 to 2016.

Structure

The IAEA is a member of the United Nations family of organizations and shares much of the UN's structures, processes, and culture. It is located at the Vienna International Centre along with other Vienna-based UN organizations.

Though it reports annually to the United Nations General Assembly and, on request, to the UN Security Council, the IAEA is not a UN specialized agency like UNESCO or the World Health Organization. Rather, the IAEA is an autonomous organization governed by its member states through a general conference, in which all member states are represented, and a 35-member board of governors. In theory, the general conference, which convenes annually, sets broad policy that guides the board of governors. In reality, power at the IAEA is concentrated in the board, both by design and evolved practice. The board comprises semi-permanent members repeatedly elected due to their importance to the peaceful uses of nuclear energy, along with non-permanent members elected for two-year terms on a regional basis. This allows every member state to be represented on the board at some point.

The board holds at least six sessions per year and may also meet in emergency situations. It considers membership applications, establishes the Agency's work program and budget, and approves all agreements with member states, safety and security standards, major infrastructure, and special projects. The board has the right to declare a state in violation of its safeguards obligations and to report it to the UN Security Council for possible enforcement action, which it has done with respect to Iraq, North Korea, and Iran.

All five of the "official" nuclear weapon states (according to the NPT) and the states most advanced in nuclear energy in each region of the world are awarded virtual permanent membership on the board. Unlike the UN Security Council, no member state has veto power. While approval of the Agency's program and budget requires a two-thirds majority, only a simple majority is required for all other matters.

Apart from its headquarters in Vienna, the Agency has regional offices in Tokyo and Toronto and research laboratories in Seibersdorf, Austria, and Monaco. The staff of the Agency, known as the secretariat, comprises approxi mately 2,560 multidisciplinary professional and support staff from more than 100 countries.

All are international civil servants recruited according to UN regulations, with consideration given to geographical (and more recently, gender) balance. The Agency is headed by a director general who is appointed by the board of governors, with the approval of the general conference, for a four-year term, which is often renewed.

Funding

The Agency is funded by assessed contributions from each member roughly according to its GDP as calculated by the UN, with significant discounts for developing countries. The Agency also relies on voluntary contributions from wealthier member states. The IAEA's total regular budget in 2022-23 was approximately $419.8 million. The Agency, along with all other UN bodies, has operated at zero real budgetary growth since 1985.

Conclusion

The IAEA aims to use international cooperation to promote the benefits of a powerful technology while also limiting the harm it can do to humanity. In this purpose, it has much in common with a potential AI regime. Because IAEA similarly deals with a highly sensitive technology that, if misused, can pose an existential threat, the Agency's experience suggests several lessons for the international governance of AI.

Do not underestimate the potential pervasiveness of a given technology. In the earliest years of the "nuclear age", it appeared that only the most sophisticated countries could pursue nuclear technology. This was soon proven false. The same is even more likely to apply to the spread of AI capabilities, where the misuse of AI could be perpetrated not only by any state but by any citizen of any state. Even if universal participation in a governance organization is unachievable at the outset, it will be important to have all the major players involved in negotiating and initiating implementation of an AI regime. Promotional efforts to achieve universality could follow, as was the case with the IAEA. Additionally, the division of states into permanent "haves" and "have nots" as in the IAEA should be avoided.

Avoid giving veto power to a single member in the quest for universality. In the case of the Comprehensive Nuclear Test Ban Treaty, negotiators wanted to lock in all states that had tested nuclear weapons in the past or could so in the future.

This gave states like India, Egypt, Iran, North Korea, and Pakistan, which are on the "essential country" list, an effective veto over entry into force of the treaty. As a result, the implementation organization for the treaty can operate only in provisional mode, which prevents activity such as on-site inspection in case of a suspected nuclear test. This situation should be avoided by an AI regime: better to have most of the key players involved than to hold out for all of them.

Incorporate industry from the beginning. In the early nuclear industry, governments were the principal drivers of technological innovation and calls for governance. In the AI industry, newly emerging expertise is confined to a relatively small number of corporations and countries. Given the difference in who is driving technological innovation and the lessons learned by IAEA, figuring out how to bring both companies and governments to the table at the earliest stages will be critical. An alternative to the IAEA model is the International Labor Organization, a "tripartite" UN body where industry and trade unions are represented along with governments.

While both nuclear radiation and AI are intangible to the average person, nuclear material is a physical artifact that must be dug out of the ground, processed, refined, enriched, shaped into the form of an explosive, mounted on a delivery vehicle, and launched. Only then does it become deadly. Obtaining weapons-grade nuclear material is a high bar to anyone contemplating using nuclear weapons. The rapid evolution of AI, its seeming malleability, and its potential for misuse by a wide variety of actors with as yet unknown effects make AI fundamentally different and potentially even harder to govern than nuclear. At the same time, the success that international organizations like the IAEA have achieved in avoiding catastrophe and encouraging collaboration between nations in the nuclear realm offers hope for achieving similar outcomes in other fields.

3.4 The Intergovernmental Panel on Climate Change (IPCC)

Authored by Diana Liverman and Youba Sokona

Purpose

The Intergovernmental Panel on Climate Change (IPCC) was established by the United Nations in 1988 to provide regular scientific assessments of climate change—including climate science, impacts and responses—that would reduce the risks and "consider uncertainties and gaps in knowledge about climate change and information needed for responses and policies".[119]

The international scientific community, and some governments, pushed for the establishment of IPCC because of the potential risks that increasing atmospheric carbon dioxide and so-called greenhouse gases produced by human activity would lead to the warming of the planet and other climatic changes that would endanger people and ecosystems. Scientists saw the need for careful review of emerging published research on climate change and options for mitigation (the reduction or removal of greenhouse gases) and adaptation (processes of adjusting the climate changes that have occurred or could occur).

The assessments published by IPCC are intended to provide policy makers, especially UN member governments, with information to develop climate policy as well as to inform the negotiations and agreements developed under the United Nations Framework Convention on Climate Change (UNFCCC).

The first set of IPCC reports was published in 1990. IPCC's first report laid out the risks of climate change and the need for international cooperation. This report played a key role in the creation of the UNFCCC, which was approved by 154 nations at the 1992 UN Rio Conference on Environment and Development Earth Summit. The UNFCCC set out to prevent "dangerous human interference with the climate system" with an initial focus on stabilizing and reducing greenhouse gas concentrations in the atmosphere.

Since then, IPCC reports written by scientific experts from around the world have been published every five to seven years, with the latest 6th Assessment cycle releasing their final reports from 2021 to 2023. Report outlines and summaries are approved by an intergovernmental panel of delegates at IPCC plenaries. Reports are usually acknowledged, discussed and accepted by the UNFCCC. The UNFCCC is the United Nations institution of political governance, whereas the IPCC is a scientific governance process.

Since its founding, the IPCC has influenced research and the UN negotiations process, as well as informing decision making by local and national governments, citizens, and the private sector. The IPCC, by identifying gaps in knowledge, has also influenced the research agendas and funding of thousands of scientists, as well as governments and foundations.

Specifically designed to inform international policy and treaties on climate change, the IPCC is, for the most part, admired and respected for providing critical scientific and technical input for the governance of climate change through the UNFCCC, governments and other international environmental treaties.

The impact of the IPCC received worldwide recognition in 2007 when it was awarded the Nobel Peace Prize jointly with Al Gore for "their efforts to build up and disseminate greater knowledge about man-made climate change, and to lay the foundations for the measures that are needed to counteract such change".

The IPCC's most significant impact on global governance has been its engagement with the UNFCCC and its decisions. Perhaps the IPCC's most influential report, requested by the UNFCCC with the 2015 Paris Agreement and published in 2018, examined potential impact of limiting global warming to 1.5°C and the pathways to achieving this goal. The report's main conclusions were that impacts of warming increased significantly from 1.5°C to 2°C, and that to have a good chance of keeping warming under 1.5°C, the world needed to cut emissions in half by 2030 and reach net zero by 2050. Countries, cities, corporations, and citizens regularly refer to this report in their climate commitments.

Despite the efforts of IPCC, and nations' efforts to agree on responses through the UNFCCC, human-generated greenhouse gas emissions have continued to increase, albeit at a slower rate. Global temperatures are the highest on record, with serious impacts on human wellbeing and the natural world. When IPCC was founded in 1988, the concentration of CO2 in the atmosphere was about 350 ppm. This has continued to increase each year, reaching 420ppm in 2023. In 1988, global

emissions of greenhouse gases were equivalent to about 38 gigatons per year. By 2022, greenhouse gas emissions had reached almost 55 gigatons. The global average temperature has risen by 1.2°C since 1880, and the 10 warmest years on record have occurred since 2010.

History

The risk that carbon dioxide emissions would warm the planet has been known for more than 50 years. Since the 1970s, key scientific papers have projected a doubling of carbon dioxide concentrations in the atmosphere associated with fossil fuel burning and deforestation and outlined potential impacts on temperatures, crop yields and sea levels.[120] The Keeling curve, which showed the steady rise of CO2 in the atmosphere at Mauna Loa, became an early emblem of a pending climate crisis.

International cooperation on climate research and applications can be traced to the establishment of the International Meteorological Organization (IMO) in 1873 to share weather data and forecasts. The formation of this organization acknowledged that weather transcends national boundaries and global observations are important. IMO later became World Meteorological Organization (WMO). In 1972, the pivotal UN Stockholm Environment confe rence included discussion of rising CO2 levels and the risks of global warming.

In 1979, the first World Climate Conference included presentations on climate change, the atmosphere as a common concern of humanity, the need for international agreements on weather modification, and the increase in carbon dioxide associated with fossil fuel use. The conference recommendations include the need for international assessments of future global climate.[121]

In the opening keynote for the 1979 World Climate Conference,[122] Bob White (of the US National Academy of Sciences Climate Research Board) made a farsighted comment:

> "You may ask, 'Why should the climate community extend its concern so far beyond scientific and technical matters into the realm of economics and social structure?' The answer is clear: Our task is to identify not just what it is that science should do, but what it is that governments should know. Unless there is a better comprehension of the chain of events and the complex interactions that take place, governmental decisions to mitigate the economic, social, and other effects of climatic impacts may very well provide the wrong remedies."

The Conference Declaration was clear about the risks of climate change driven by human activities:

> Nevertheless, we can say with some confidence that the burning of fossil fuels, deforestation, and changes of land use have increased the amount of carbon dioxide in the atmosphere by about 15% during the last century and it is at present increasing by about 4% per year. It is likely that an increase will continue in the future. Carbon dioxide plays a fundamental role in determining the temperature of the Earth's atmosphere, and it appears plausible that an increased amount of carbon dioxide in the atmosphere can contribute to a gradual warming of the lower atmosphere, especially at high latitudes. Patterns of change would be likely to affect the distribution of temperature, rainfall and other meteorological parameters, but the details of the changes are still poorly understood.

The origins of the IPCC can be traced to a series of events and meetings in the 1980s. A UN workshop in Villach, Austria, in 1985 convened experts from 29 countries to assess the impacts of rising CO2.[123] The Villach conference statement called for periodic assessments of the state of scientific understanding and its practical implications, and proposed a global convention, perhaps inspired by the successful negotiation of the 1985 Vienna Convention on the ozone layer.

In 1987, the influential Brundtland Commission highlighted the Villach report, and the World Meteorological Congress recommended that there be periodic

assessments of climate risks under the overall guidance of governments. The WMO had also initiated a series of assessments focused on the risks of atmospheric ozone depletion with reports in 1985, 1988, and 1989.[124] Bob Watson of NASA chaired these assessments, which included many scientists who later became IPCC authors. The ozone assessments underpinned the 1987 United Nations Montreal Protocol on Substances that Deplete the Ozone Layer, presaging the 1992 UNFCCC.

In 1988, 300-plus scientists and policy makers gathered in Toronto, Canada, for the "Conference on the Changing Atmosphere". Those gathered called for the establishment of an Intergovernmental Panel on Climate Change and a comprehensive global convention to protect the atmosphere. They argued for a 20% reduction in carbon dioxide levels by 2005. The Toronto conference received widespread media attention—in the US, the media was simultaneously responding to a heatwave and NASA scientist Jim Hansen's recent congressional testimony about the serious risks of global warming. High-level political attention in Europe included a major speech by British Prime Minister Margaret Thatcher about global warming as a massive experiment on the planet. In November 1988, the UN established the IPCC at a session of the WMO, and it was then endorsed at the UN General Assembly.

The creation of IPCC was based on international network of scientists who saw the systemic risk and need for response to anthropogenic climate change. It became an intergovernmental organization partly because countries such as the US wanted some control over the assessments.

While it was initially challenging to develop IPCC's principles, the organization's core principles have endured. These include that assessments should:

- Be based on scientific expertise and a balanced and comprehensive analysis of the state of knowledge;
- Be based, to the extent possible, on peer-reviewed scientific literature;
- Go through review by other scientists and by governments;
- Seek consensus; and
- Be policy relevant but not policy prescriptive.[125]

In some cases, governments and climate skeptics saw any discussion of solutions as political and policy prescriptive, and challenged the scientists in

plenary sessions and the media. A principle of consensus has endured, with scientists working incredibly hard to agree upon their conclusions and attaching careful statements about confidence, uncertainty, likelihood and lines of evidence. This effort at consensus is also evident in the conduct of government delegates at the Summary for Policy Makers report approval sessions.

Evolution

The first (1990) and second (1995) assessment reports of the IPCC, as well as specific technical reports on assessing emissions, regional impacts, sea level rise, and potential climate scenarios, were particularly important in identifying trends in different greenhouse gases—notably adding methane to the well-known trends in carbon dioxide—as well as the human activities that produced them, especially fossil fuel use and land use change.[126] IPCC also synthesized what was known about climate trends and developed glossaries that defined key terms such as mitigation and adaptation. IPCC assessed the results of complex models that analyze the future impacts of greenhouse gases on global temperature and the earth system as well as a set of socioeconomic scenarios to project the emissions associated with different demographic, technological, and policy futures.

Just as the suite of material released by IPCC has evolved over time, IPCC connections to the UNFCCC have also evolved. At COP3 in 1997, the UNFCCC, recognizing the risks identified by IPCC, adopted the Kyoto Protocol, in which developed countries made binding commitments to reduce emissions and carbon trading mechanisms were established to increase flexibility. The IPCC Task Force on National Greenhouse Gas Inventories supported national reporting to the convention on emissions.

The third assessment report, released in 2001, was notable for its focus on vulnerability to climate change—particularly the disproportionate impacts on polar regions, small islands, and Africa—as well as the importance of adaptation. The literature cited and the conclusions of the assessment have underpinned the negotiation positions of UNFCCC, national groups of common interests such as the Association of Small Island States (AOSIS), and the Climate Vulnerable Forum.

The fourth assessment (2007) laid the ground for the target of limiting warming to 2°C. Debates at COP 15 in Copenhagen and the fifth assessment report (2013/14) underpinned the pivotal COP 21 Paris Agreement in 2015. UNFCCC

has also made several requests for special reports from the IPCC, including one on regional vulnerability (1997), technology transfer (2000), and land use and forestry (2000).

One of the contentious issues for the UNFCCC climate negotiations has been identifying what level of global temperature rise constitutes "dangerous" interference with the climate system. The EU had identified 2°C as a potential target, while vulnerable countries have called for lower targets. This debate emerged in the tense negotiations at COP 15 over the Paris Agreement, which included the goal of keeping global temperature rise below 2°C and eventually 1.5°C. The agreement included a request to the IPCC to assess these goals.

The impact of the 1.5°C report, published in 2018, went far beyond interested scientists and the climate negotiations.[127] Climate activists, including youth movements, took to the streets to put pressure on policy makers. Major corporations made pledges to halve emissions by 2030 and reach net zero by 2050. Congress in the US and other governments around the world held hearings and made emission reduction pledges. The EU revised policies to reduce emissions by 55% below 1990 levels by 2030 and to be climate neutral by 2050. After US President Joe Biden was elected in 2020, his administration aligned with the 1.5°C target by aiming to reduce emissions 52% below 2005 levels by 2030.

Achievements and challenges

IPCC's significant scientific achievements include reducing uncertainty in understanding how greenhouse gas emissions drive climate change and compiling the evidence that changes are occurring and can be attributed to human activities. In each successive report, the scientific evidence for observed climate change and future projections has become more robust and confident. While the 1990 report highlighted many uncertainties about ongoing and future climate change, the most recent report concludes, with high confidence, that "human activities, principally through emissions of greenhouse gases, have unequivocally caused global warming".

Despite the large research literature now available for assessment by IPCC, there are significant gaps in the science and literature required to cover the full scope of issues and make confidence statements and to serve the needs of international and local climate governance.

Detailed analyses of national responsibility for emissions are critical to negotiations about who should reduce emissions and who should pay or reductions by others or compensate them for impacts. Because greenhouse gases remain in the atmosphere for long periods, historical responsibilities are an important element of negotiations. IPCC avoids pointing fingers at specific nations or regions because it will cause problems with member countries. IPCC also avoids assessing the responsibility of specific companies— such as fossil fuel majors—and aggregates emissions by sector. The UNFCCC convention included recognition of differing responsibilities and capacities of developed and developing countries in "common but differentiated" responsibilities and respective capabilities. The recent focus on the relationships between climate and sustainable development is a step towards acknowledging vastly different levels of development.

Some impacts of climate change are understudied in the literature and cannot be confidently assessed by IPCC. These include, for example, impacts on certain regions, on the manufacturing and service sectors, on workers, on culture, and on supply chains and trade. Earlier gaps in research on climate impacts on health, cities, and food systems are now better addressed. Another related change is that scientific literature on climate change impacts is often reliant on case studies rather than comparative or aggregated assessments, in part because local governments such as cities often benefit most from case studies focused on their particular risks and solutions.

Assessment of the economic costs and impacts of climate change are limited, with an overreliance on integrated and aggregated models as well as controversial assumptions about discounting and non-market values. During the approval plenary for the 1.5°C report, several developing countries expressed disappointment at the lack of quantitative or economic data on costs and climate impacts on their regions and economies. Additionally, a number of fossil fuel-producing countries wanted information on how energy transitions could damage their economies. These issues are of heightened importance given the new UNFCCC negotiating track on "loss and damage".

Geoengineering solutions, which focus on removing greenhouse gases from the atmosphere or reducing solar radiation inputs, have not been fully analyzed by IPCC or addressed by global governance. On the other hand, greenhouse gas removal through land use, especially through protecting and restoring forests, has been a strong focus of IPCC from the beginning. In recent years, assessment has broadened to look at the role of other land uses, including coastal ecosystems, for

capturing carbon. Technological solutions that involve capturing CO2 at power stations or from the air have been assessed by IPCC as not yet economically feasible or scalable. Solar radiation management, which would compensate for warming by reflecting incoming solar radiation through putting sulfur or other particles into the atmosphere, is now briefly mentioned in IPCC reports. However, solar radiation management is not mentioned as a mitigation option, and reports include cautions about the risks of unanticipated and unequal consequences of implementing or halting the technologies.

Governance

IPCC operates within the United Nations system under the auspices of the WMO and the United Nations Environment Programme (UNEP). As an intergovernmental institution, it is managed by the IPCC Plenary, which discusses plans and budgets, approves the Summaries for Policy Makers of major reports, and elects the scientific members of the IPCC Bureau. The Plenary meets at least once per year in different host countries and is attended by government delegates, some scientists, and some observer organizations.

Government delegates vote on key issues, with one vote per country and a tradition of consensus approval of reports. Most countries are represented by government-designated "focal points" who may be from foreign affairs or environment departments, or from meteorological research institutes. Representatives may or may not be well informed about climate issues, especially if they are from weather bureaus, but are increasingly trained in diplomatic negotiations and are often laser focused on the wording of the IPCC summaries. IPCC does not have a legal personality or engage in treaty making.

Organizations with observer status can send someone to plenaries to attend but not speak (in principle, although they will often interact with scientists and delegates outside the room). These include other international and regional organizations (such as the World Bank, InterAmerican Institute for Global Change Research, and African Union Commission) and NGOs such as the C40 Cities climate leadership group, Greenpeace, Oxfam, and the Stockholm Environment Institute.

Initial funding for IPCC was set up in 1989 through the IPCC Trust fund, with contributions from WMO, UNEP and member countries. The trust fund has accumulated a balance of about $20 million, with the largest country contributions

since 1989 coming from France, Germany, Japan, the UK, and the US. In 2022, contributions totaled about $2.5 million. The annual budget for IPCC is about $8.5 million, with about $4 million for expenses incurred by the Secretariat, including publications, IT, and communications; $2.8 million for the plenary and governing meetings; and $1 million for author meetings.

IPCC is staffed by a small, 14-person secretariat based at WMO in Geneva and includes legal, logistical, and communications staff. Scientific work, including major assessment reports, is managed by a Bureau of 34 scientists elected by the member countries and includes the chair and vice-chairs of the overall IPCC and its Working Groups. There are three major scientific Working Groups:

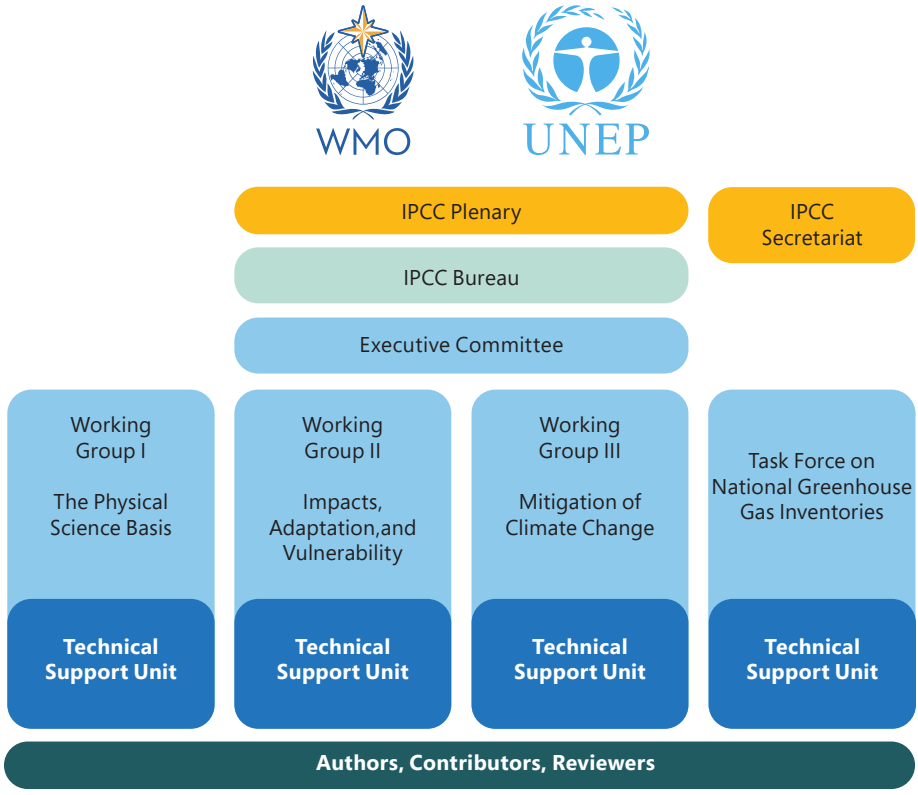
- Working Group I: Assesses the physical science of the climate system and climate change. This includes the understanding of climate processes, observations of climate change, and climate modeling.
- Working Group II: Focuses on impacts, adaptation, and vulnerability. This group assesses the impacts of climate change on natural and human systems, the capacity of societies to adapt, and options for reducing vulnerability.
- Working Group III: Addresses the mitigation of climate change. This includes options for reducing greenhouse gas emissions, economic and technological issues, and activities that remove greenhouse gases from the atmosphere.

Each Working Group is coordinated by a small Technical Support Unit that organizes logistics, reviews, and other everyday needs of the working groups. Some of the co-chairs from the Global South receive support for technical assistance in their work. The IPCC also has a Task Force on National Greenhouse Gas Inventories (TFI), which develops and assesses methods for inventorying emissions.

The UNFCCC provides the main conduit for IPCC to influence global governance, with the technical arm of the UNFCCC serving as the primary formal link between IPCC and UNFCCC. The annual UNFCCC Conference of Parties (COP), hosted by a different country each year, is the main venue for discussion, assessment of progress, and negotiation, with a more technical meeting held at the UNFCCC secretariat in Bonn each summer.

IPCC is also asked to present at every COP. The UNFCCC regularly expresses appreciation for IPCC's reports, invites their presentation at COPs, and provides some funding to IPCC.

IPCC operations and governance



Source: IPCC

The IPCC report process

Since the first comprehensive assessment in 1990, the working groups write separate reports—though special reports, such as 1.5°C, have been developed jointly. The final synthesis report is also written by scientists selected from all three working groups. The preparation and releaseof the major assessment reports is staged with the release of the Climate Science (WG1) report several months before the Impacts (WG2) and then the Mitigation (WG3) reports. The Working Groups try to connect and coordinate their messages, but this is not always successful.

Each report starts with a scoping meeting of experts nominated by member governments and the IPCC Bureau; experts prepare an outline, which becomes the approved outline for the report. This is followed by a call to nominate authors, mostly through national governments but also through the Bureau. Authors are selected based on their expertise related to the approved outline, but increasingly to ensure a balance of gender, geography, and disciplines and to ensure there are some contributors with prior experience as IPCC authors, especially in selecting who will lead the chapters as Coordinating Lead Authors. Each chapter of a report has Coordinating Lead Authors (2-3) and 10-20 Lead authors, with some scientists invited to write small sections as Contributing Authors. Each chapter has a chapter scientist, most often a younger scholar and two review editors who have prior experience with IPCC and will ensure that review comments are addressed.

The evolution of the IPCC has seen six rounds of assessment reports so far, beginning with the 1990 First Assessment and with the latest Sixth Assessment—which ran from 2015 to 2023. Each round includes a set of comprehensive assessment reports produced by the working groups as well as special reports on specific topics. Special reports for the Sixth Assessment included reports on Climate Change and Land, the Ocean and Cryosphere, and Global Warming of 1.5°C.

Authors for each Working Group then convene several times to draft the report with oversight from the Working Group chairs and vice chairs who are members of the IPCC Bureau. Each Working Group has a Technical Support Unit with a small number of paid staff to coordinate the report preparation. Pre-COVID, meetings lasted about one week, with about four meetings for each report. At the initial meetings, the chapter group decides how to implement their chapter outline and starts to compile the relevant peer reviewed literature, asking authors to start writing in their areas of expertise. In some cases, the Coordinating Authors dominate the action; other chapters work more collectively. A first draft usually emerges about half way through the process and is made available for expert review by fellow scientists. Almost anyone can apply to be a reviewer and will be given access to the draft, although they are expected to have some expertise and not to leak the report. Some chapters receive thousands of review comments.

Chapters are then rewritten in response to review, and a second order draft is prepared and opened for government review. A final draft is then prepared, responding to reviews, updating literature, and polishing conclusions, and submitted to the Bureau. Around this time, a subset of authors, led by the Working

Group Co-chairs, starts to prepare the Summary for Policy Makers (SPM). This is the most important part of every IPCC report and is what receives most media, political and public attention. It summarizes the conclusions of the report and must be approved, sentence by sentence, by member governments at an IPCC plenary. Approval of the SPM is eagerly awaited by the media and science community, and IPCC now employs a sophisticated communications team to manage press inquiries and train scientists to talk about the reports.

Challenges in governance for IPCC

There have been important governance challenges for the IPCC over the years. These include:

1. **Political interference with IPCC processes:** Politics has inevitably entered into the operations of the IPCC. The process of electing the bureau has become very political, with countries vying to have their scientists elected and doing side deals to gain support for their candidates. Requests for input from the UNFCCC often reflect tensions between countries, such as decisions about temperature targets, who is most vulnerable, and the technologies and funding of responses. Government reviews of report drafts and approval sessions of the Summary for Policy Makers also tend to reflect international politics. Some countries prepare extensive comments from several government agencies that are mostly constructive but are clearly trying to limit conclusions. Some governments object to any discussion of equity and justice, claiming it to be normative rather than objectively scientific. In some cases, conclusions are toned down through scientists' self-censorship or government changes to the Summary for Policy Makers.

2. **Efforts to delegitimize IPCC reports:** Every IPCC report receives criticism from climate skeptics. Perhaps the most notable example of this was "Climategate" in 2009, where the hacking of servers at the University of East Anglia resulted in the release of hundreds of emails between IPCC authors as they prepared reports. Critics interpreted several emails to suggest scientists were biased in their assessments of temperature trends and impacts. Despite robust responses, the scandal partly derailed COP 15 in Copenhagen.[128] Such attacks have made IPCC careful to avoid leaks of report drafts, though they still occur, and to check every statement and line of evidence. IPCC governance has also developed a protocol to investigate and address alleged errors in reports as well as potential conflicts of interest.

3. **Author burn-out:** IPCC authors are not paid by IPCC. While a few may be released from their regular job duties (usually those working for government research groups), the majority are volunteers working in their spare time. Assessment reports take up to four years to prepare, with special reports operating on shorter timelines. Additionally, the amount of scientific literature that needs to be reviewed has grown exponentially since the first report. Authors of the latest Working Group II report on impacts cited an overwhelming 34,000 articles and responded to 62,000 review comments. The IPCC plenaries, where the Summary for Policy Makers from reports are approved, involve negotiators and scientists working around the clock for days. IPCC authors also experience great frustration and even climate grief when they see the results of their reports attacked, ignored, or producing inadequate policy responses.

4. **Bias in author and bureau selection:** Author selection has been criticized for overlooking women, people from the Global South, representatives from the private sector, and NGO and social science experts, as well as indigenous and younger scholars. For example, early reports included very few female authors, even when there were senior women available. Even now, women comprise only one-third of the authors. Early reports also lacked representation from developing countries. Even as representation is broadened to include more female or developing world scholars, surveys find that individuals from these communities often do not feel as though their voices are heard. An official IPCC gender task force is attempting to improve the situation.

5. **Weak policy responses:** Despite progress in areas such as electric vehicles (EVs) and renewables, temperatures on land are now 1.74°C warmer than they were in 1850. IPCC previously estimated that emissions needed to drop about 5% per year from 2018 to 2030 to keep global temperature increase under 1.5°C. IPCC also estimated that our remaining carbon budget was about 500 gigatons (GT). Now, given the delay in action, we need to achieve a steep drop in emissions every year to 2030, and our remaining carbon budget is only 250 GT remaining. Net-zero goals and carbon neutrality promises are unrealistic if they rely on technologies that are not yet economically feasible or scalable. Recent research found that most countries, corporations, and cities that have made net zero pledges are making assumptions about negative emissions or carbon offsets without viable strategies.[129] This lack of progress has been linked to global geopolitics such as the Ukraine war, the voluntary nature of the Paris agreement, continued subsidies for fossil fuels, climate obstruction by

fossil fuel companies, the rebound of aviation and consumption post COVID, countries with high historical emissions refusing to make deeper cuts, and extreme weather/AC demand.

6. **Communication challenges:** In its efforts to provide comprehensive assessments of climate change, the IPCC has produced ever longer assessment reports. Whereas the first set of reports was about 1,000 pages, the latest set reached over 10,000 pages. Efforts to better communicate results of the assessments include a carefully curated web site, shorter Summaries for Policy Makers (35-50 pages for each Working Group), translations into various UN languages (Arabic, Chinese, French, Russian, and Spanish), a set of brief headline statements, technical summaries, fact sheets, downloadable figures and slide presentations, and social media campaigns. Nevertheless, there are many calls for much shorter reports, more frequent brief updates on the state of science, combining of assessments into a single working group, greater focus on key policy questions, and improved graphics. IPCC itself, and the scholarly community, have proposed reforms of IPCC processes over the years.[130] These include reducing political influence, increasing diversity of voices, writing much shorter reports, merging or reorganizing the working groups, softening the emphasis on consensus, reducing reliance on computer models, including more social sciences and humanities, adding more stakeholder authors, and giving more attention to local knowledge and solutions.

How IPCC has influenced other governance processes

Global environmental governance has seen extensive cooperation on a set of important treaties and conventions, including the UN conventions on long-range air pollution (1979), the ozone layer (1985), and desertification (1994). Most of these assessments have been underpinned by scientific assessments with similar processes to IPCC.

Perhaps the most important siblings to the IPCC and UNFCCC are the UN Convention on Biological Diversity and the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES), both of which aim to protect biodiversity.[131] IPBES was modeled on IPCC. While supported by UNEP, IPBES is not an official intergovernmental body of the UN and has a stronger focus on local knowledge than IPCC. In 2021, IPCC and IPBES issued a joint report on biodiversity and climate change.

There have been calls for IPCC-like scientific assessments for health, AI, and geoengineering, but these sometimes idealize or misunderstand the IPCC, the challenges it has faced, and its important relationship with the UNFCCC.

Conclusion

The IPCC represents an important model for the global governance of systemic risks that also seeks to inform policy. IPCC emerged due to growing concern about climate change promoted by an epistemic community of scientists who volunteered to assess peer-reviewed literature. IPCC reports have influenced international negotiations as well as actions and awareness of national and local governments, non-governmental organizations, businesses, and the general public, and have shaped scientific research agendas. But IPCC is not without its critics, limitations, and gaps in knowledge. New science-based assessment proposals should pay close attention to these challenges.

Proposals for an IPCC-type assessment process for AI should take into account key aspects of IPCC governance, including the way intergovernmental status both benefits and politicizes IPCC assessments and policy impact, the vital connection between IPCC and the UN climate convention (UNFCCC), and the challenges of including all countries and stakeholders in the assessments.[132] Does the most important knowledge on the risks and possibilities of AI exist in an open, peer-reviewed literature? How can private and defense sector insights become part of such assessments without conflict of interest, competitive issues, or security risks? Should a scientific assessment of AI be closely linked to UN or other multilateral agreements on AI safety? Could such an assessment rely on consensus between authors and unanimous government agreement to approve reports?

Above all, a key lesson from the experience of IPCC is that, despite decades of warnings about climate change, action has been delayed and limited, and the risks are still existential and immediate. It is vitally important that an intergovernmental organization related to AI not only deepens knowledge but also hastens solutions, rather than underestimate risks and distract from action.

3.5 The Bank for International Settlements (BIS), Basel, the Financial Stability Board (FSB), and the Financial Action Task Force (FATF)

Authored by Christina Parajon Skinner

Introduction

After World War II, many Western states expressed commitment to global economic cooperation as a means of ensuring lasting peace.[133] By the 1980s, the arena of international finance came to be increasingly governed by soft-law institutions, which consist of networks of financial regulators. This chapter explains the architecture of that system with an aim to provide lessons, inspirations, and cautionary tales for a possible global framework to govern—more specifically, set safety standards around—the risks presented by artificial intelligence (AI).

Today, a number of international financial regulatory bodies set international standards for globally active banks and other financial institutions. These bodies identify possible risks that these institutions could present to the global economy (or, conversely, risks presented to these institutions' safety and soundness), and share information across borders. Most of these organizations arose in response to economic crises or gaps in public international law in the realm of financial supervision and risk.

These bodies include, most notably, the Bank for International Settlements (BIS), which formally hosts both the Basel Committee for Banking Supervision (Basel or BCBS) and the Financial Stability Board (FSB). These institutions mainly focus on risks relevant to banks and systemically important market based credit intermediators, like money market funds.[134]

The Financial Action Task Force (FATF), meanwhile, is a global standard-setting body with a focus that's risk-specific rather than industry-specific. FATF seeks to address the global challenges presented by money laundering, the financing

of terrorism, and the proliferation of weapons of mass destruction. Often, the financial system is at the center of these problems, but increasingly money is laundered for these illicit ends through crypto assets, real estate, art, and other non-financial companies.

Purpose

These organizations exist to serve four main purposes:

1. **To address the risk of regulatory arbitrage.** Certain risks are global in nature and thus cannot be mitigated by any one (or handful of) jurisdictions. A porous system allows for what is known as regulatory arbitrage, whereby risky or unlawful behaviors simply shift from more regulated to relatively laxer geographies. These international institutions thus endeavor to address the possibility of arbitrage by working toward the harmonization of basic standards.

2. **To minimize informational blind-spots.** High quality and complete information is essential to early or preventative action. However, without cooperation among national regulators, blind-spots emerge. Accordingly, much of these institutions' work is geared toward sharing knowledge, information, and best practices across jurisdictions.

3. **To advance international comity.** In a crisis, cooperation among regulators is crucial. Establishing sound working relationships builds good will and trust across staff and can smooth crisis-time interventions. Accordingly, these institutions form and maintain networks between national regulators and supervisors from a wide range of jurisdictions.

4. **To apply moral suasion.** Given the risk of regulatory arbitrage, lax enforcement or non-compliance by a handful of jurisdictions can undermine the efforts of others to plug gaps in legal and supervisory frameworks. Accordingly, these bodies have developed a system of peer monitoring, review, and feedback that is ultimately meant to pressure jurisdictions into cooperating and publicly "naming-and-shaming" them if they do not.

Over the past century, this system for the global governance of finance has evolved to take on an increasing number of tasks and promulgate an intricate set of standards. The lessons of this experiment are, however, somewhat mixed. On

the one hand, the system has worked relatively well at coordinating principles and ideas. On the other hand, it suffers from perennial challenges to its legitimacy—which will only grow along with the scope and mission of this framework—a lack of transparency, and the tendency to occasionally distort outcomes on the national level.

The most notable legal feature of the BIS, Basel, FSB, and FATF is that they have no formal legal status. Unlike formal international economic institutions that are constituted and governed by treaty—like the WTO, the World Bank, and the International Monetary Fund—these international financial regulatory organizations exist only pursuant to "soft law".[135] They exist because regulators decided to agree, amongst themselves, to form these networks and cross border association.[136] While these regulators have come to agreement about chartering provisions and governance procedures, these organizations' existence has not been authorized by relevant national legislatures. It bears emphasis that these central bankers and other bank regulators are not, themselves, elected or democratically responsive. The soft law nature of these networks of regulators poses interesting and often overlooked question about the force of their prescriptions.

Each of these institutions formally acknowledges their informal, soft law status. The Basel Committee, for example, states in its charter that "Its conclusions do not have, and were never intended to have, legal force. Rather, it formulates broad supervisory standards and guidelines and recommends statements of best practice in the expectation that individual authorities will take steps to implement them through detailed arrangements – statutory or otherwise – which are best suited to their own national systems."[137] The FSB's charter, in similar spirit, acknowledges in its Article 23: "This Mandate is not intended to create any legal rights or obligations."[138] Using identical language, Article 48 of the FATF Charter notes that "This Mandate is not intended to create any legal rights or obligations."[139] Of course it must be stated so—short of a formally ratified treaty, no institution that exists in international law can impose binding obligations on national sovereign states.

Still, their respective charters also require members to commit to implementing their standards. Basel committee members "agree to implement fully Basel standards for their internationally active banks. These standards constitute minimum requirements and BCBS members may decide to go beyond them."[140] FATF members must likewise "endorse and implement the FATF Recommendations for combating money laundering and the financing of terrorism

and proliferation, using where appropriate guidance and other policy endorsed by the FATF."[141] The Basel and FATF charters also require members to commit to peer review process, which will be discussed in further depth.

History

In some form or another, war motivated international cooperation in the banking space. The first of these efforts was the creation of the BIS, formed in 1930 at the Hague Conference. Its initial job was the complicated task of settling, in as neutral a fashion as possible, the reparation payments that were imposed on Germany after World War I. Specifically, the BIS managed the collection and then administration and distribution of the annuities payable as reparations. Later, it would facilitate the issuance of German bonds through the Dawes and Young programs.[142]

The story of the BIS is one of evolution and adaption. After the cessation of reparation payments in the 1930s, the BIS evolved its role to promote technical cooperation between central banks, including on matters involving reserve management, foreign exchange transactions, gold deposits, and swap facilities; it also convened and provided a forum for meetings of central bankers.[143] After the abandonment of the Bretton Woods Agreement—the essence of the international gold standard—the BIS evolved yet again to focus principally on its coordinating role and to a lesser known extent, the provision of financial service for central banks.[144]

The BIS is owned by national central banks; 63 different central banks own its shares. By accepting currency and gold deposits, and investing through proceeds to earn a profit, its balance sheet resembles a national central bank.[145] As such, the BIS functions somewhat like an international central bank, albeit without the ability to set anything like global monetary policy. It does, however, indirectly influence national central banking policy by hosting two distinct international regulatory institutions, Basel and the FSB.

The Basel Committee was established by the central bank Governors of the G10 countries at the end of 1974.[146] Though it was not formed in the heat of war, per se, an increasing number of disturbances in the international currency and banking markets prompted reflection on how best to close gaps between the supervision and regulation among increasingly internationally active banks.[147] Today, 45

institutions from 28 different jurisdictions are members of the Basel Committee. Members are generally central banks or authorities for prudential (that is, safety and soundness) supervision in their country.[148]

The FSB was established much later, in the wake of the global financial crisis of 2008, though it had its origins in a separate, now defunct, body, the Financial Stability Forum (FSF). The FSF was launched in 1999 in response to the Asian financial crisis, with a goal of analyzing risks that, if materialized, could propagate and adversely affect wide swaths of regional or global economies—known as "systemic risk".[149]

After the 2008 crisis, the G20[150] met in Cannes and decided to transform the FSF into the FSB and bolster its mandate and capacity.[151] In particular, it was agreed that the FSB would gain an "enduring organisation footing, strengthening its coordination role vis-à-vis other standard-setting bodies on policy development and implementation monitoring, and reconstitution of the FSB's Steering Committee."[152] Although the G20 technically sits atop the FSB, the BIS hosts the FSB physically and provides for its Secretariat. Perhaps for this reason, the FSB has focused its attention on many central banking prudential matters and tends to work closely with Basel.

FATF is a reaction to the global war on drugs and terror. It was initially developed by the G7 in the 1980s, in response to the drug trade being financed by money laundered through the global banking system.[153] For the leaders in these nations, it was "clear that there needed to be a coordinated response. No country could fight money laundering on its own."[154] Money laundering is a global problem that is difficult to mitigate. The UN Office on Drugs and Crime estimates that the amount of money laundered globally each year is about 2-5% of global GDP—according to IMF estimates, this is about $1.6-$4 trillion annually.[155] The internationally active banks that form correspondent banking networks remain a key battleground of governments' fight against it. Accordingly, financial policy makers convene globally to set international standards of best practices for combatting money laundering, corruption, and terrorist financing through the FATF. FATF now has nearly 40 members.

Evolution

In domestic law, all administrative agencies—which include financial regulators—have mandates and responsibilities set out in statute. These

organizations also have mandates and objectives set out in their charters. Again, however, these mandates have been developed by the institutions' members—not by any domestic or supranational legislature. These mandates have been framed quite broadly, which has over time supported the expansion of these institutions' scope and functions.

The FSB's mandate is to "promote global financial stability by coordinating the development of regulatory, supervisory and other financial sector policies and conducts outreach to non-member countries."[156] To accomplish that objective, the FSB has two main functions. The first is a standard setting role. It has, since 2012, developed standards and principles that have cross-sectoral implications for multiple jurisdictions. Examples include Key Attributes of Effective Resolution Regimes for Financial Institutions[157] and a set of policy recommendations for dealing with the risks presented by non-bank credit intermediation.[158]

The FSB also engages in monitoring or early warning work. It studies what it identifies as emerging financial stability risk and publishes research and working papers that direct attention to certain areas. It is difficult to say what the impact of this work product is. The FSB also has some ability to drive forward collective problem-solving in areas of high concern to the G20. The FSB's ongoing work to tackle the efficiency of cross-border payments is a current case in point.[159]

The output of this type of work can take the form of something like a "roadmap" for national jurisdictions to follow toward the collective goal. These outputs are less formal than standards—which carry the expectation of implementation—yet more concrete than papers identifying emergent risks.

The Basel committee's mandate is also set out in its charter, which stipulates: "The BCBS is the primary global standard setter for the prudential regulation of banks and provides a forum for cooperation on banking supervisory matters. Its mandate is to strengthen the regulation, supervision and practices of banks worldwide with the purpose of enhancing financial stability."[160] Its function in pursuit of that goal as evolved over time. In its early days, Basel was principally focused on coordinating supervisory standards. Today it is best known for its setting of internationally agreed capital adequacy standards of banks in various Basel Accords.

The Third Basel Accord, adopted in 2010 following the 2008 financial crisis, has involved multiple inter-linking layers, including capital adequacy standards, liquidity standards, stable funding standards, and new guidance for supervisory

stress testing.[161] Most advanced economies have over the years been diligent in implementing the Basel agreements; the US is currently in the last phases of implementing Basel III, colloquially referred to as the Basel "endgame."

FATF's objective is to set anti-money laundering (AML) and combatting the financing of terrorism (CFT) standards. More specifically:

> The objectives of the FATF are to set standards and to promote effective implementation of legal, regulatory and operational measures for combating money laundering, terrorist financing and other related threats to the integrity of the international financial system. In collaboration with other international stakeholders, the FATF also works to identify national-level vulnerabilities with the aim of protecting the international financial system from misuse.[162]

Accordingly, the chief function of FATF is the development of its recommendations, standards for the effective detection of money laundering and other forms of illicit finance. Naturally, the recommendations have iterated over the years to account for new ways that criminals launder money; most recently, the recommendations have been updated to address digital assets.[163]

Historically, FATF has faced more difficulty than Basel in securing widespread compliance. Part of this is because money laundering happens in a wider range of jurisdictions than those that are home to large, internationally active banks. This leads to broad differences in capacity and willpower to enforce. Even to secure the minimum level of buy-in, in order to coordinate the recommendations, FATF has had to face the political and practical reality that each country's circumstances differ, and participation will only be maximized with the offer of flexibility. As such, FATF subscribes to a risk-based approach, in which "countries, as well as private sector, identify, assess and understand the risks they are exposed to and focus their resources on areas where the risks are highest."[164]

Governance

The global governance framework for international finance is highly unique in the history of public international law and international economic coordination. These institutions have considerable authority to set rules of the game for private institutions—most of which are tremendously costly and restrictive to their business models—but, as noted, no formal basis in law and little political accountability. This distinct legal and institutional design is both a strength and weakness of this governance model.

Governance at each institution is generally set up to enable these bodies to "remain a flexible, responsive, member-driven, multi-institutional and multidisciplinary institution".[165] Nominally, the G20 sits atop of most of it, although the funding comes mostly from the central banks.

The FSB's main body is the "Plenary" of the entire membership. The Plenary is populated by senior policymakers from ministries of finance, central banks, and supervisory and regulatory bodies form the G20 countries. It is led by a Chair that rotates among senior officials from the members for a three year term.[166] However, most of the work is done at the steering committee and standing committee levels. The Plenary appoints members to both the steering committee, which drives forward the plenary agenda, and the standing committees, which move forward the various workstreams of the FSB. Standing committees are led by senior level officials from the member states.[167] The FSB also has a Secretariat, which is directed by the Secretary General who is appointed by the Plenary. Technically, these employees contract directly with BIS.

The Basel Committee reports to the central bank governors and heads of bank supervisors from the G10—the group is referred to as the Group of Central Bank Governors and Heads of Supervision (GHOS). GHOS acts like an oversight body and appoints the Basel Committee Chair, who serves as the external face of the Committee. Basel also has a Secretariat funded and hosted by the BIS.

The FATF follows a similar governance structure, with a Plenary that is responsible for agenda-setting, which in turn creates working group populated with members chosen based on the Plenary President's recommendation. These working groups and steering groups are responsible for "taking forward, in consultation with the plenary, any other work necessary for the FATF to fulfill its mandate".[168] The IMF and World Bank also play a significant role in solidifying

FATF's work by conducting country assessments and providing technical assistance and capacity building in the anti-money laundering and counter-terrorism financing spaces.

The primary implication of these bodies' soft law status is that the standards they create do not have any binding force in law. They must be implemented into domestic law pursuant to the jurisdiction's usual process for promulgating rules and other kinds of public law. In the United States, this means that whatever standards central bankers or Treasury officials might agree to internationally, they must be re-written to be specific to the US financial and economic system and will only become binding once successfully finalized through the notice-and-comment rulemaking process required by the Administrative Procedure Act.[169]

This leads to questions of enforcement. Even treaty-based public international law struggles to enforce its rules—and those institutions have the political support, and in theory might, of the state behind them. How, then, do these soft-law, non-binding, network-focused institutions compel compliance with their standards, recommendations, and roadmaps?

The short answer to this question is that formally they cannot. Still, each of these three institutions engages in soft enforcement in the form of peer review. The FSB uses two types of peer reviews: thematic reviews and country reviews.[170] Thematic reviews consider how effectively members are implementing FSB standards.[171] Thematic reviews can also address other areas important for global financial stability where international standards or policies do not yet exist. The reviews are meant to "encourage" implementation and make recommendations to members about how they might fill in identified gaps.[172]

Country reviews, on the other hand, are connected to the IMF-World Bank Financial Sector Assessment Program (FSAP) and Reports on the Observance of Standards and Codes (ROSCs) recommendations on financial regulation and supervision. Beyond the FSAP compliance, these country reviews can also focus on regulatory, supervisory, or other financial sector policy issues "that are timely and topical for the jurisdiction itself and for the broader FSB membership".[173]

The Basel Committee takes a similar-in-spirit approach. It monitors implementation of its standards through its Regulatory Assessment Programme (RCAP), established in 2012. RCAP has two main elements: monitoring and assessment. By compiling information periodically submitted by members, BCBS maintains a monitoring dashboard that is publicly available; assessments,

meanwhile, involve the constitution of a cross-jurisdictional evaluative team and result in the formal publication of a graded report card of sorts.[174]

Like the FSB, the FATF uses peer reviews, called Mutual Evaluations, to diagnose problems and evaluate implementation of FATF Recommendations.[175] Mutual Evaluations are framed around both effectiveness and technical compliance. An effectiveness assessment entails a visit from an assessment team—the assessed country will have to demonstrate evidence that their measures are working.[176] The technical compliance aspect of a Mutual Evaluation entails the assessed country providing "information on the laws, regulations and any other legal instruments it has in place to combat money laundering and the financing of terrorism and proliferation".[177] These evaluations are performed regularly, and all reports are published by the FATF.[178]

Conclusion

The standards set by the financial governance organizations discussed in this chapter have had considerable influence in domestic law. Still, the reader should recognize that this governance paradigm has its limits, many of which have not yet been fully tested.

The first of these concerns the lack of political accountability. These institutions are not responsive to their members' domestic legislatures and yet their standards often ultimately become imported into law.[179] Occasionally, this oddity grabs the attention of lawmakers and shines an unpleasant light on their work.[180] The more stringent the standard, the more likely it will be to raise questions about the legitimacy of the soft-law process.

Although these bodies do engage in public consultation, and sometimes include academics and private sector representatives in their working groups, ultimately, they alone decide the content of any standards set. It is difficult for the public to know which individuals, exactly, populate these working level groups. So, the public cannot know these participants' interests, objectives, or incentives. Relatedly, because these institutions' inner workings are opaque, it is easy for them to become captured by the special interests of outside groups.

Ultimately, there are plenty of reasons why soft law regimes are more agile, innovative, and adaptable than formal institutions. For that reason, a realist might say that any ability to secure coordination is superior to none. But because broad-based domestic political buy-in is essential to the long-term viability of this tenuous framework, they can only go so far in pushing outside the bounds of the public and lawmakers' reasonable expectations of transparency and accountability.

4

Looking Back to Look Ahead

Governing AI is a vast, multidimensional, and iterative project. AI will affect how we live and work; how every major industry operates; how governments serve their citizens; how criminals act nefariously; and how conflicts are waged. Around the world, private sector companies, governments, civil society, and academia will contribute to understanding AI's opportunities and risks and defining and ensuring implementation of effective guardrails.

But as was clear from the knowledge shared with us by our group of experts, AI is not the first domain to require complex and ever evolving global governance. Take civil aviation, nuclear power, and global capital flows. At the dawn of the 20th century, early flight experimentation set the stage for decades of change to war, commerce, and culture. Enrico Fermi's 1934 discovery that neutrons could split atoms entwined devastating weapons with the emergence of a potentially pivotal global energy industry. And our modern financial system was disrupted by the Great Depression and two world wars before it was then revived, enabling innovation while creating global systemic risk.

Civil aviation, nuclear power, and global capital flows have prompted governance by industry, domestic authorities, and international institutions. The balance across each layer of regulation has varied, with public-private partnerships playing a stronger role for civil aviation and global capital flows and international institutions being granted more authority with nuclear power. These variations reflect differences in the technologies and risks being governed as well as the historic moments in which these governance systems emerged and began to evolve.

Today, governments, industry, and civil society are actively advancing industry standards, domestic regulation, and international governance for AI, creating an opportunity to build in interoperability and cohesion across a broad constellation of initiatives from the start.

As the international community commits to building a more robust system of AI governance, we see value in developing frameworks that help reinforce a coherent direction and coordinated action among a proliferation of tremendously useful initiatives at the global and domestic levels. We see value in reflecting on other historical moments that have called upon global leaders to create durable institutions as well as the ways in which time and circumstance have tested them, motivated their evolution, and demonstrated their impact. And we see value in continued exchange among diverse experts, including those we've welcomed the opportunity to learn from through their contributions to our thinking and this publication. Given the rapid advancements in technology that we're witnessing and the momentum of AI policy discussions, learning from other domains can help ground efforts to build out a framework and agenda for international AI governance.

Ultimately, though, we need to use this context to look ahead. Recognizing the many efforts at play and many interests at stake—and the resulting imperative of collaboration—is at the foundation of this AI governance project. We need collaboration to help weave together mutually reinforcing initiatives, reducing their permeability by reinforcing their seams—or to build from a common cloth, adding local color and details as we bring together the resources and capabilities more inherent to a global system. Learning from the decades of experience that have defined our modern, highly interconnected world, international initiatives and institutions are likely to play a critical role in facilitating this collaboration. Institutions can bring focus to new or evolving functions and grow expertise needed to take on more complex and multi-faceted governance projects, enabling collaboration toward a shared vision even in complex geopolitical environments.

International institutional purposes and functions will also be key to realizing the three international AI governance outcomes that Chapter One proposed: globally significant risk governance, regulatory interoperability, and inclusive progress.

How we act—and toward what outcomes and with what impact—matters not just for AI technology but also and much more importantly for the social, environmental, economic, and political futures that are interwoven with it. The consequences of AI governance thus reverberate for organizations, communities, and people everywhere. They call upon us to be inclusive and collaborative, representing the many interests at stake and the many efforts that will ultimately accrue to effective, interoperable global AI governance.

Key lessons for AI from existing domains of global governance

- Domains presenting global challenges and opportunities require global governance.
- Policymaking is more effective if grounded in scientific or technical research and a deep understanding of the challenges to be addressed.
- Effective governance frameworks define core functions and desired outcomes.
- For global frameworks to succeed, proactive and strategic leaders play a critical role in building and broadening support.
- Multistakeholder collaboration at the technical and political levels is important to develop robust global standards and allow for rapid response to emergent risks.
- Successful governance systems evolve over time, with old and new institutions taking on a mix of interconnected functions and objectives.

5

Recent Multilateral Developments

This section provides an overview of the variety of developments in the area of AI governance at the international level over the last 12 months.

UN initiatives

In the past year, there have been several UN initiatives to address AI, some of which stemmed from proposals of the **UN Secretary-General** and others driven by UN organizations and processes.

In March 2024, the **UN General Assembly** adopted a resolution to promote safe, secure, and trustworthy AI systems for sustainable development. It was adopted by consensus and co-sponsored by more than 120 countries. The resolution highlighted the need to respect, protect and promote human rights in the design, development, deployment, and use of AI, and also recognized the potential of AI to accelerate and enable progress towards reaching the UN Sustainable Development Goals.

Separately, UN Member States will develop a **UN Global Digital Compact (GDC)** to be adopted as part of the Summit of the Future in September 2024. The GDC is expected to "outline shared principles for an open, free and secure digital future for all", including on AI. Throughout 2023, AI was featured in many of the GDC consultation stakeholder submissions (including from Microsoft), as well as the Secretary-General's own GDC policy proposal. The co-facilitators of the GDC process noted in an Issues Paper that AI is emerging as a key issue for the GDC. The process has included input from a wide array of stakeholders including UN member states, industry, civil society, academia, non-governmental organizations, and youth.

A key input to the GDC and the Summit of the Future will be the work of a new **UN High-Level Advisory Body on AI**. The 39-member multistakeholder and interdisciplinary body (which includes Microsoft's Chief Responsible AI Officer,

Natasha Crampton, in her personal capacity) published an interim report in December, and will make final recommendations in summer 2024 in three areas: international governance of AI, understanding AI's risks and challenges, and opportunities to leverage AI to deliver the UN Sustainable Development Goals.

The **United Nations Educational, Scientific and Cultural Organization (UNESCO)** continues its work to support implementation of its 2021 Recommendation on the Ethics of AI. Its February 2024 Global Forum on the Ethics of AI focused on the changing landscape of AI governance. In 2023, it launched a UNESCO Business Council for Ethics of AI to help ensure that AI is developed and utilized in a manner that respects human rights and upholds ethical standards. The AI Business Council (of which Microsoft is a co-chair) is committed to strengthening technical capacities in ethics and AI, designing and implementing the Ethical Impact Assessment tool mandated by the UNESCO Recommendation, and contributing to the development of regional regulations.

The **UN Office of the High Commissioner for Human Rights (OHCHR)** and its B-Tech Community of Practice launched a Generative AI Human Rights Due Diligence Project in May 2023. The project looks at how the UN Guiding Principles on Business and Human Rights (UNGPs) can guide more effective understanding, mitigation, and governance of the risks of generative AI.

The **International Telecommunication Union (ITU)** works with stakeholders to build a common understanding of the capabilities of AI technologies to facilitate the trusted, safe, and inclusive development of AI technologies, and equitable access to their benefits. Its AI for Good platform promotes AI to advance health, climate, gender, inclusive prosperity, sustainable infrastructure, and other global development priorities. In July, the ITU's annual AI for Good Global Summit included discussions about the need for guardrails and global AI governance frameworks. The 2024 Summit in May 2024 will for the first time include an AI Governance Day.

In July 2023, the **UN Security Council** convened a session on "Artificial Intelligence: Opportunities and Risks for International Peace and Security". The session, led by the UK during its presidency of the Security Council, was the Council's first-ever discussion on AI.

AI was also a major topic of discussion at the annual meeting of the **UN Internet Governance Forum (IGF)** in October 2023. The overview of the topics discussed included views on several elements of AI policy: global cooperation,

governance, human rights and development, and generative AI. An IGF Policy Network on AI also made recommendations in a report entitled *Strengthening multistakeholder approach to global AI governance, protecting the environment and human rights in the era of generative AI*.

In October 2023, the United Nations Institute for Disarmament Research (UNIDIR) launched a report on "AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures". This report creates a taxonomy of the risks of AI in the context of international peace and security and provides a comprehensive overview of these risks and how they may impact global security. Multistakeholder discussions on Confidence-Building Measures for AI are expected to commence in early 2024.

In October 2023, the **World Health Organization (WHO)** released a new publication listing key regulatory considerations on AI for health. It outlines key principles that governments and regulatory authorities can follow to develop new guidance or adapt existing guidance on AI. It emphasizes the importance of establishing the safety and effectiveness of AI systems, rapidly making appropriate systems available to those who need them, and fostering dialogue among stakeholders, including developers, regulators, manufacturers, health workers, and patients.

In October 2023, the **United Nations Third Committee** (focused on Social, Humanitarian and Cultural Issues) in New York started discussions on a draft resolution on the "Promotion and protection of human rights in the context of digital technologies". The draft notes that AI can contribute to the promotion and protection of human rights and has the potential to transform governments and societies, economic sectors, and the world of work. It calls upon the private sector and all relevant stakeholders to ensure that respect for human rights is incorporated into the conception, design, development, deployment, operation, use, evaluation, and regulation of all new and emerging digital technologies.

Intergovernmental initiatives

Alongside these developments at the global level through UN bodies, there are also discussions and initiatives taken by smaller groups of governments.

In 2023, the **G7** produced a Hiroshima AI Process Comprehensive Policy Framework, which included a Code of Conduct for organizations developing advanced AI systems and Guiding Principles for all AI actors. The 2024 G7 Digital Ministerial Declaration committed to working with the OECD on tools and mechanisms to monitor application of the Code of Conduct, and to broaden the involvement of key partners and organizations.

The **G20** 2023 Leaders Declaration in September reaffirmed a commitment to the G20 AI Principles (2019) and the pursuit of a "pro-innovation regulatory/ governance approach that maximizes the benefits and takes into account the risks associated with the use of AI" and promotes "responsible AI for achieving SDGs".

In November 2023, the UK hosted the **AI Safety Summit**, attended by 27 governments, the EU, UN, and tech companies, including Microsoft, DeepMind, Meta, and OpenAI. The summit had a number of outcomes: the Bletchley Declaration signed by all attending governments and the EU, a commitment to a "State of the Science" report on the capabilities and risks of frontier AI, a partnership between the UK and US AI Safety Institutes, and a Chair's statement on safety. Additional AI safety summits will take place in South Korea and France in the coming year.

OECD, GPAI, and other initiatives

It is also important to consider the work of the OECD and others where governments work with stakeholders to incorporate technical expertise into policy analysis to advance thinking on various aspects of AI governance.

The 38-country **Organisation for Economic Cooperation and Development (OECD)** continued its wide range of work on AI—a Working Party on AI Governance leads work on AI policy while a separate AI Network of Experts provide technical, academic, and business expert input. Outputs in 2023 included Initial policy considerations for generative AI and a report on AI and Jobs that explored future skills needs. The OECD also updated the definition of an AI system

within its 2019 AI Principles to reflect the emergence of generative AI; a full review of the OECD AI Principles will be undertaken in the first half of 2024.

The **Global Partnership on AI**, a multistakeholder initiative which provides a mechanism for sharing multidisciplinary research and identifying key issues among AI practitioners, released a policy brief on Generative AI, Jobs, and Policy Response and a report on AI Foundation Models & Detection Mechanisms.

UNESCO, the OECD, GPAI, and other partner organizations launched a Global Challenge to Build Trust in the Age of Generative AI. Over the next two years, it will surface and test innovative ideas to promote trust and counter the spread of disinformation.

References/Citations

1 See Chapter 2.

2 "Global leaders agree to launch first international network of AI Safety Institutes to boost cooperation of AI," UK Government, May 21, 2024.

3 "International Scientific Report on the Safety of Advanced AI – Interim Report," UK Government.

4 Brad Smith, "Microsoft's AI Access Principles: Our commitments to promote innovation and competition in the new AI economy," Microsoft On the Issues, February 26, 2024.

5 Millions of people began using natural language and generative AI systems to create text, code, and audio-visual media, helping people achieve better outcomes in education, energy, healthcare, and more. "NTNU helps K-12 students master English faster with innovative learning platform powered by Azure OpenAI Service," Microsoft Customer Stories, April 19, 2023; "EDP: powering the global energy transition using Microsoft Intelligent Data Platform, AI and IoT technology," Microsoft Customer Stories, June 30, 2023; "Healthcare for All with Kry using Azure Open AI Service," Microsoft Customer Stories, October 24, 2023; "When everyone speaks the same language: Boehringer Ingelheim speeds up knowledge sharing with Azure OpenAI Service," Microsoft Customer Stories, October 24, 2023; "Providence uses Azure OpenAI Service to decrease clinician burnout and expedite patient care," Microsoft Customer Stories, November 10, 2023.

6 Interim Report: Governing AI for Humanity, United Nations, December 2023: 22. Microsoft's Chief Responsible AI Officer is Natasha Crampton, included among other members listed here.

7 "The Global AI Talent Tracker 2.0," MacroPolo.

8 Describing "impermeability" as one of five guiding principles. Ian Bremmer and Mustafa Suleyman, "Building Blocks for AI Governance," International Monetary Fund, December 2023.

9 For example, ICAO develops standards that span a wide range of safety and security issues in the civil aviation sector. This includes setting out standardized specifications for critical pieces of safety equipment, such as the collision avoidance systems, through pilot and crew licensing, evaluation of aircraft worthiness, and airport regulation.

10 "Mutual Recognition Agreements," European Commission.

11 *See* Chapter 3.

12 Id.

13 Discussing these "problem" questions in connection with the purposes for international institutions Keohane defined. Robert Keohane, After Hegemony: Cooperation and Discord in the World Political Economy, Second Edition (New Jersey: Princeton University Press, 2005).

14 "'State of the Science' Report to Understand Capabilities and Risks of Frontier AI: Statement by the Chair, 2 November 2023," Gov.UK – AI Safety Summit, Nov. 2, 2023.

15 "International Scientific Report."

16 "Global Digital Compact Second Revision," United Nations, June 26, 2024; "Global Digital Compact Third Revision," United Nations, July 11, 2024.

17 Trager, et al., "International Governance of Civilian AI: A Jurisdictional Certification Approach," September 11, 2023.

18 "Global leaders agree to launch first international network of AI Safety Institutes."

19 Compute refers to both computational infrastructure—the hardware necessary to develop and deploy an AI system—and computational power, the performance of the AI chip commonly measured in integer or floating-point operations. Highly capable AI training requires large amounts of AI chips performing a large number of computations. *See also* other works on compute governance: Lennart Heim and Leonie Koessler, "Training Compute Thresholds: Features and Functions in AI Governance," May 17, 2024; Konstantin Pilz and Lennart Heim, "Compute at Scale: A Broad Investigation into the Data Center Industry," Nov. 5, 2023.

20 "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," The White House, October 30, 2023; "Artificial Intelligence Act," European Parliament, March 13, 2024.

21 *See* Heim, Fist, Egan et al. Governing Through the Cloud: The Intermediary Role of Compute Providers in AI Regulation, 26 March 2024.

22 The organization providing the model, for example, could be required to demonstrate that they have a suitable risk management plan in place. This could include a process for performing risk identification and mitigation, implementing cybersecurity requirements, and monitoring for and responding to significant incidents.

23 "123 Agreements for Peaceful Cooperation," National Nuclear Security administration, US Department of Energy.

24 Chris Welsch, "Assisted by AI, a workforce of bees tracks pollution and boosts biodiversity," Microsoft News Centre Europe, September 18, 2023.

25 "OECD AI Principles Overview," OECD.

26 "G20 Ministerial Statement on Trade and Digital Economy," Ministry of Foreign Affairs of Japan; "The state of implementation of the OECD AI Principles four years on," OECD, October 27, 2023.

27 "Recommendation on the Ethics of Artificial Intelligence," UNESCO, May 16, 2023.

28 "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems," G7 Hiroshima Summit 2023.

29 "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development," United Nations General Assembly, March 11, 2024.

30 "TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management," National Institute of Standards and Technology, December 1, 2022.

31 "Hiroshima Process International Code of Conduct."

32 "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems."

33 In doing so, it will provide an AI equivalent to ISO/IEC 27001 and ISO/IEC 27701, foundational standards in the cybersecurity and privacy domains.

34 "Information technology — Artificial intelligence — AI system impact assessment," International Organization for Standardization.

35 "G7 Industry, Technology and Digital Ministerial Meeting Ministerial Declaration," G7 Italia, March 15, 2024.

36 "Apulia G7 Leaders' Communique," G7 Italia 2024.

37 "Reporting Framework for the International Code of Conduct for Organizations Developing Advanced AI Systems: Pilot Phase," OECD.AI Policy Observatory.

38 "GPAI and OECD unite to advance coordinated international efforts for trustworthy AI," OECD, July 3, 2024.

39 "Member countries of the Hiroshima AI Process Friends Group,".

40 Describing existing guidance and international instruments with which the draft guidelines will dock in, *see* OECD Due Diligence Guidance for Responsible Business Conduct, OECD, 2018.

41 *See* Recital 127: "In line with Union commitments under the World Trade Organization Agreement on Technical Barriers to Trade, it is adequate to facilitate the mutual recognition of conformity assessment results produced by competent conformity assessment bodies, independent of the territory in which they are established, provided that those conformity assessment bodies established under the law of a third country meet the applicable requirements of this Regulation and the Union has concluded an agreement to that extent. In this context, the Commission should actively explore possible international instruments for that purpose and in particular pursue the conclusion of mutual recognition agreements with third countries."

42 Brad Smith and Melanie Nakagawa, "Accelerating Sustainability with AI: A Playbook," Microsoft On the Issues, November 16, 2023.

43 "AI for science," Microsoft Research.

44 Elliott Smith, "AI may hold a key to the preservation of the Amazon rainforest," Microsoft Source LATAM, September 6, 2023.

45 "Research Forum Episode 2: Transforming health care and the natural sciences, AI and society, and the evolution of foundational AI technologies," Microsoft Research, March 6, 2024.

46 Tie-Yan Liu and Tao Qin, "GHDDI and Microsoft Research use AI technology to achieve significant progress in discovering new drugs to treat global infectious diseases," Microsoft Research, January 16, 2024.

47 "The 17 Goals," United Nations.

48 Inclusiveness is one of Microsoft's foundational AI principles. "Principles and Approach," Microsoft AI/Responsible AI.

49 Smith, "Microsoft's Access Principles."

50 Many international organizations continue to focus and make progress on broadband infrastructure, including the International Telecommunications Union (ITU) and the UN Educational, Scientific and Cultural Organization (UNESCO). "Universal Broadband Connectivity," International Telecommunication Union. Microsoft also recognizes this imperative and is actively working to bridge gaps in these areas, including via our Airband Initiative, through which we've connected 51 million people to high-speed internet since 2017 and set the goal to connect 250 million people, including 100 million Africans, by 2025. "The ITU/UNESCO Broadband Commission for Sustainable Development," Broadband Commission.

51 Basic research, perhaps especially at universities, is of fundamental importance to countries' economic and strategic success; the past few decades have seen huge swaths of research in almost every field propelled by growing compute resources and data science. "Governing AI: A Blueprint for the Future," Microsoft, May 25, 2023.

52 Last October, the US AI Executive Order tasked the National Science Foundation with launching a pilot program implementing the National AI Research Resource (NAIRR), pursuing the infrastructure, governance mechanisms, and user interfaces to make available computations, data, models, and training resources to support AI-related research and development. "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." In November, the UK announced that it will build and connect two supercomputers, giving researchers access to resources with more than 30 times the capacity of the UK's current largest public AI computing tools. "Technology Secretary announces investment boost making British AI supercomputing 30 times more powerful," UK Government, November 1, 2023. In April, Canada announced a $2.4 billion investment, including to build and provide access to AI infrastructure for local researchers, start-ups, and scale-ups; help bring new technologies to market and support deployment among critical sectors and small businesses; support workers; strengthen enforcement of a proposed law; and create a new Canadian AI Safety Institute. "Securing Canada's AI advantage," Prime Minister of Canada, April 7, 2024.

53 Through Accelerate Foundation Models Research (AFMR), we've selected 125 new projects from 75 institutions across 13 countries. AFMR is a research grant program through which we aim to facilitate interdisciplinary research on aligning AI with human goals, values, and preferences; improving human interactions via sociotechnical research; and accelerating scientistic discovery in natural sciences. After managing a pilot phase that launched earlier in 2023, we expanded the program and have now selected 125 new projects from 75 institutions across 13 countries. The focus of our first open call for proposals was on aligning AI systems with human goals and preferences; advancing beneficial applications of AI; and accelerating scientific discovery in the natural and life sciences. As we continue to expand the breadth of our reach with academic partnerships, we will also continue to expand the depth of our research, including in areas like AI evaluation and measurement. "Microsoft's AI Safety Policies," Microsoft On the Issues, October 26, 2023.

54 "Microsoft announces a $5 billion investment in computing capacity and capability to help Australia seize the AI era," Microsoft Australia News Centre, October 24, 2023; Brad Smith, "Our investment in AI infrastructure, skills and security to boost the UK's AI potential," Microsoft On the Issues, November 30, 2023; Smith, "Microsoft's AI Access Principles"; "Microsoft to invest US$2.9 billion in AI and cloud infrastructure in Japan while boosting the nation's skills, research and cybersecurity," Microsoft Stories Asia, April 10, 2024.

55 Smith, "Microsoft's AI Access Principles."

56 MCDF highlighted that its investment aims to enhance regional "research, development, and application capabilities of AI solutions through the HPC network," starting in Chile and the Dominican Republic. The MCDF grant will cover a series of feasibility studies, including an HPC supply and demand analysis, a roadmap for constructing HPC centers in Chile and the Dominican Republic, and a proposal based on technical, legal, environmental, institutional, and other assessments. "MCDF Grant to Open Door to Ai Computing Network in Chile and Dominican Republic," Multilateral Cooperation Center for Development Finance, September 21, 2023.

57 "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems."

58 Miriam Brady, "Microsoft Launches New AI Skills Training and Resources as part of Skill for Jobs Initiative," Microsoft Nonprofit Community Blog, October 24, 2023.

59 "Artificial intelligence and the Futures of Learning," UNESCO, September 12, 2023.

60 "Business Council for Ethics of AI," UNESCO.

61 The Initiative puts forward new, free coursework developed with LinkedIn, including the first Professional Certificate on Generative AI in the online learning market, as well as a new open global grant challenge in coordination with data.org to uncover new ways of training workers on generative AI; it also advances greater access to free digital learning events and resources for everyone to improve their AI fluency. There are also resources dedicated to schools and educators and pathways for specific roles, including engineering. "AI Skills," Microsoft Corporate Social Responsibility; "Microsoft Azure AI Fundamentals: Generative AI," Microsoft Build.

62 Smith, "Microsoft's AI Access Principles."

63 "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems."

64 Id.

65 AI for Good consists of an online program and annual in-person summit, co-convened with Switzerland and 40 other UN agencies. The ITU also facilitates Focus Groups on AI for Natural Disaster Management, for Digital Agriculture, and for Health; these groups are helping assist with data collection and handling, improving the precision and sustainability of farming techniques, and evaluating AI-based health methods. "Artificial Intelligence," International Telecommunication Union.

66 "AI For Good Lab," Microsoft Research

67 "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems," G7 Hiroshima Summit 2023; "Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI," The White House.

68 "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems."

69 "G7 consensus reached on advancing AI for sustainable development," United Nations Development Programme, March 15, 2024.

70 "Rebalancing the Global AI Landscape," ICAIN.

71 While we joined events hosted by other organizations, these reflections solely represent Microsoft's takeaways from those conversations and do not represent the views of any organizations that hosted events.

72 "Global Perspectives: Responsible AI Fellowship," Stimson. *See also* Natasha Crampton and Brian Finlay, "Advancing AI responsibly," Microsoft.

73 "Interim Report: Governing AI for Humanity," United Nations AI Advisory Body, December 2023, (describing seven overlapping functions of AI governance).

74 "The AI Dialogues, Interim note #1: Is international AI governance achievable?" Renaissance Numérique, June 2024, (describing seven overlapping AI governance functions).

75 "Artificial Intelligence," ITU, (including work on AI for Good and capacity development); "Artificial Intelligence," UNESCO, (including work on the ethics, digital capacities building, and AI and inclusion).

76 "GPAI and OECD unite to advance coordinated international efforts for trustworthy AI."

77 "The Common Criteria," Common Criteria; "Developments in the field of information and communications in the context of international security," United Nations; Brad Smith, "The need for a Digital Geneva Convention," Feb. 14, 2017; "Paris Call for Trust and Security in Cyberspace,".

78 "AI Action Summit," Elysée.

79 David Brazier, "The Atlantic Charter: Revitalizing the Spirit of the Founding United Nations Over Seventy Years Past," United Nations Chronicle.

80 Notably, Roosevelt also understood the need for domestic political buy in. To build up political support for this new institution, he brought both administration and elected officials in negotiations, and sought to ensure Congress was supportive throughout the endeavor. For more on this point, *see* "The United States and the Founding of the United Nations, August 1941 – October 1945," Office of the Historian, Bureau of Public Affairs, US Department of State.

81 IO numbers drawn from Jon C.W. Pevehouse, Timothy Nordstron, Roseanne W McManus, and Anne Spencer Jamison, "Tracking Organizations in the World: The Correlates of War IGO Version 3.0 datasets." Journal of Peace Research 57, no. 3 (2020).

82 For a discussion of US leadership and Western commitment to a rule-based order, *see* G. John Ikenberry. 2019. After Victory: Institutions, Strategic Restraint, & The Rebuilding of Order After Major Wars (New Edition). Princeton: Princeton University Press, and G. John Ikenberry. 2009. "Liberal Internationalism 3.0: America and the Dilemmas of Liberal World." Perspectives of Politics 7(1): 71-87. On the rise of global norms, *see* G. John Ikenberry, After Victory: Institutions, Strategic Restraint, & The Rebuilding of Order After Major Wars (New Edition), (Princeton: Princeton University Press, 2019) and G. John Ikenberry, "Liberal Internationalism 3.0: America and the Dilemmas of Liberal World," Perspectives of Politics 7, no. 1 (2009): 71-87. On the rise of global norms, *see* Michael Barnett and Martha Finnemore, Rules for the World: International Organizations and Global Politics, (New York: Cornell University Press, 2004). On the link between the number of countries and the number of IOs, *see* Pevehouse, Nordstrom, McManus, and Jamison, "Tracking Organizations in the World."

83 Informal IO numbers drawn from Felicity Vabulas and Duncan Snidal, "Cooperation under autonomy: Building and analysing the Informal Intergovernmental Organizations 2.0," Journal of Peace Research 54, no. 4 (2021): 859-869.

84 Julia C. Morse, The Bankers' Blacklist: Unofficial Market Enforcement and the Global Fight Against Illicit Financing. (New York: Cornell University Press, 2022).

85 *See* Tyler Pratt, "Deference and Hierarchy in International Regime Complexes," International Organization 72,3 (2018): 561-590 for work on patterns of institutional deference within regime complexes.

86 Julia C. Morse and Robert O. Keohane, "Contested Multilateralism." The Review of International Organizations 9 (2014): 385-412.

87 In Keohane's book After Hegemony, he describes these three advantages as reducing information asymmetries, establishing legal liability, and reducing transaction costs. I have modified these terms to make them more accessible to a general audience. For a more detailed description, *see* Robert O. Keohane, After Hegemony. (Princeton, NJ: Princeton University Press, 1984).

88 Declaration of the Summit on Financial Markets and the World Economy.

89 Morse, The Banker's Blacklist.

90 "Factsheet on DPRK Nuclear Safeguards," IAEA.

91 Anne Marie Slaughter, A New World Order. (Princeton, NJ: Princeton University Press, 2005).

92 Barnett and Finnemore, Rules for the World.

93 Vabulas and Snidal, "Cooperation under autonomy."

94 *See* ICAO, "Chicago Conference Introduction," International Civil Aviation Organization.

95 *See* "About ICAO," International Civil Aviation Organization.

96 Convention on International Civil Aviation, "Preamble," International Civil Aviation Organization.

97 *See* FAA Advisory Circular No. 20-151C.

98 *See* 81 Fed. Reg. 96,572 (Dec. 30, 2016).

99 *See* "Department of Transportation Pipeline of Hazardous Materials Safety Administration," ICAO.

100 *See* "Safety Management System Frequently Asked Questions," Federal Aviation Administration.

101 *See* "Details on FAA Noise Levels, Stages, and Phaseouts," Federal Aviation Administration.

102 Ken German, "2 Years After Being Grounded, the Boeing 737 Max is Flying Again," CNET, June 19, 2021.

103 Allison Lampert, "Russia Loses U.N. Aviation Council Seat in Rebuke," Reuters, October 1, 2022.

104 "Reducing Emissions From Aviation," European Commission.

105 "CORSIA Fact Sheet," International Air Transport Association.

106 Rafael Schvartzman, "EU ETS Reform Destabilizes International Consensus for Aviation Carbon Reductions," International Air Transport Association, April 18, 2023.

107 Allison Martell and Allison Lampert, "China Denounces U.N. Aviation Emissions Plan in Blow to Industry Efforts," Reuters, September 24, 2019.

108 Id.

109 "ICAO Annex Forward: SARPs Definition and Actions," International Civil Aviation Organization.

110 Id.

111 "U.S. – European Union Safety Agreement," Federal Aviation Administration.

112 "Federal Aviation Administration Returns Mexico to Highest Aviation Safety Status," Federal Aviation Administration, September 14, 2023.

113 Id.

114 CERN's stated mission is to: perform world-class research in fundamental physics; provide a unique range of particle accelerator facilities that enable research at the forefront of human knowledge, in an environmentally responsible and sustainable way; unite people from all over the world to push the frontiers of science and technology, for the benefit of all; and train new generations of physicists, engineers, and technicians, and engage all citizens in research and in the values of science.

115 For instance, the UK was chosen over Germany to house the Joint European Torus (JET) due to a hijacking in Mogadishu. Europe was chosen to house the International Thermonuclear Experimental Reactor (ITER) rather than Japan because it offered to make a much larger contribution.

116 *Access:* CERN has benefited from being next door to an international airport, whereas arguably the International Thermonuclear Experimental Reactor (ITER) has suffered from being an hour's drive from Marseille airport; openness to visitors: The host should be able to provide access to visiting scientists, although the host country generally reserves the right to deny access on good grounds; accommodation: If ample short-term accommodation is not available locally, centers that anticipate large numbers of visitors often construct hostels, thereby allowing visitors to make best use of their time and facilitating collaboration; and schooling: If the organization employs significant numbers of staff, it may be necessary to provide access to education in various languages. This is available in some cities, such as Geneva and London, which have large international populations. In other cases, special schools have been built, e.g., close to ITER, where teaching is available in six languages.

117 This author served as the scientific adviser to this External Review Committee. In the early 1990s, in the run-up to the approval of the Large Hadron Collider (LHC), which relies on what was then a very novel design of superconducting magnets, the author set up an external review of the design in order to reassure the Council, although they had not asked for such a review.

118 This paper uses the term global governance to refer to international treaties, organizations, arrangements and procedures agreed between governments to bring order and predictability to a particular realm of human activity. Although this might also be termed international regulation, it differs from national regulation. Governments have political and legal jurisdiction over their people and territory, not least a monopoly on the use of force. International organizations do not have such characteristics. While they may monitor and verify state compliance, they are only able to enforce their decisions in extreme circumstances through the intercession of the United Nations Security Council. There is no standing international police force. At the international level then, the term governance is preferable to regulation as it suggests a collective, collaborative endeavour to establish a normative umbrella over national behaviour that involves nurturing and promoting norms, standards and recommendations, codes of conduct, best practice and incentives for compliance, such as economic support and technical assistance. Monitoring and verification are increasingly possible, not least due to technological advances, but enforcement is always confronted by the doctrine of state sovereignty.

119 "Organization History," Intergovernmental Panel on Climate Change.

120 For example: Rasool, S. I, and S.H. Schneider, "Atmospheric carbon dioxide and aerosols: Effects of large increases on global climate," Science 173, no. 3992 (1971): 138-141; Schneider, SH, "On the carbon dioxide–climate confusion" Journal of Atmospheric Sciences 32, no. 11 (1975): 2060-2066; Hansen, J, et al, "Climate impact of increasing atmospheric carbon dioxide," Science 213, no. 4511 (1981): 957-966. Even earlier work by Eunice Foote (1856), John Tyndall (1860s) and Svante Arrhenius (1896) identified the potential of the greenhouse effect.

121 As a Masters student at the University of Toronto Diana Liverman helped Canadian climatologist Ken Hare, one of the organizers, prepare for the conference and then studied for her PhD under climate scientist Steve Schneider who was instrumental in creation of IPCC.

122 Proceedings of the World Climate Conference: A Conference of Experts on Climate and Mankind. World Meoteorlogical Organization 537 (1979).

123 "International Assessment of the Role of Carbon Dioxide and of Other Greenhouse Gases in Climate Variations and Associated Impacts," World Meteorological Organisation," (Villach, Austria, 1985) 537.

124 "Scientific Assesment Panel," UN Environment Program.

125 The Changing Atmosphere: Implications for Global Security. Conference Proceedings. World Meteorlogical Organization 710 (1988).

126 "IPCC Procedures," Intergovernmental Panel on Climate Change.

127 All IPCC reports are available online at this link.

128 Boykoff, Maxwell, and Olivia Pearman, "Now or Never: How Media Coverage of the IPCC Special Report on 1.5 C Shaped Climate-Action Deadlines," One Earth 1, no. 3 (2018): 285–88. Doran, Rouven, Charles A. Ogunbode, Gisela Böhm, and Thea Gregersen, "Exposure to and Learning from the IPCC Special Report on 1.5 C Global Warming, and Public Support for Climate Protests and Mitigation Policies," Npj Climate Action 2 (2023). Livingston, Jasmine E., and Markku Rummukainen, "Taking Science by Surprise: The Knowledge Politics of the IPCC Special Report on 1.5 Degrees," Environmental Science & Policy 112 (2020): 10–16. Ogunbode, Charles A., Rouven Doran, and Gisela Böhm, "Exposure to the IPCC Special Report on 1.5 °C Global Warming Is Linked to Perceived Threat and Increased Concern about Climate Change," Climatic Change 158, no.3–4 (2020): 361–75.

129 Anderegg, William RL, and Gregory R. Goldsmith, "Public Interest in Climate Change over the Past Decade and the Effects of the 'Climategate' Media Event," Environmental Research Letters 9, no. 5 (2014): 054005. Maibach, Edward, Anthony Leiserowitz, Sara Cobb, Michael Shank, Kim M. Cobb, and Jay Gulledge, "The Legacy of Climategate: Undermining or Revitalizing Climate Science and Policy?" WIREs Climate Change 3, no. 3 (2012): 289–95. Shapiro, Harold T, Roseanne Diab, Carlos Henrique de Brito Cruz, Maureen Cropper, Jingyun Fang, Louise O Fresco, Syukuro Manabe, et al., "Climate Change Assessments Review of the Processes and Procedures of the IPCC: Committee to Review the Intergovernmental Panel on InterAcademy Council," InterAcademy Council (2010): 103. Emails from our colleagues were among those whose emails were released and we spent days reading every email released, working with university attorneys, and explaining that comments were innocuous and consistent with peer review.

130 Allen, Myles R., Pierre Friedlingstein, Cécile A.J. Girardin, Stuart Jenkins, Yadvinder Malhi, Eli Mitchell-Larson, Glen P. Peters, and Lavanya Rajamani, "Net Zero: Science, Origins, and Implications," Annual Review of Environment and Resources 47, no. 1 (2022): 849–87. Hale, Thomas, Stephen M. Smith, Richard Black, Kate Cullen, Byron Fay, John Lang, and Saba Mahmood, "Assessing the Rapidly-Emerging Landscape of Net Zero Targets." Climate Policy 22, no. 1 (2022): 18–29.

131 Anderegg, William RL, and Gregory R. Goldsmith, "Public Interest in Climate Change over the Past Decade and the Efects of the 'Climategate' Media Event," Environmental Research Letters 9, no. 5 (2014): 054005. Maibach, Edward, Anthony Leiserowitz, Sara Cobb, Michael Shank, Kim M. Cobb, and Jay Gulledge, "The Legacy of Climategate: Undermining or Revitalizing Climate Science and Policy?" WIREs Climate Change 3, no. 3 (2012): 289–95. Shapiro, Harold T, Roseanne Diab, Carlos Henrique de Brito Cruz, Maureen Cropper, Jingyun Fang, Louise O Fresco, Syukuro Manabe, et al., "Climate Change Assessments Review of the Processes and Procedures of the IPCC: Committee to Review the Intergovernmental Panel on InterAcademy Council," InterAcademy Council (2010): 103. Emails from our colleagues were among those whose emails were released and we spent days reading every email released, working with university attorneys, and explaining that comments were innocuous and consistent with peer review.

132 Allen, Myles R., Pierre Friedlingstein, Cécile A.J. Girardin, Stuart Jenkins, Yadvinder Malhi, Eli Mitchell-Larson, Glen P. Peters, and Lavanya Rajamani, "Net Zero: Science, Origins, and Implications," Annual Review of Environment and Resources 47, no. 1 (2022): 849–87. Hale, Thomas, Stephen M. Smith, Richard Black, Kate Cullen, Byron Fay, John Lang, and Saba Mahmood, "Assessing the Rapidly-Emerging Landscape of Net Zero Targets." Climate Policy 22, no. 1 (2022): 18–29.

133 These include the General Agreement on Trade and Tariffs in 1947, the Bretton Woods Agreement 1944, and later the establishment of the World Trade Organization ("WTO") in 1995.

134 In the capital markets space, the International Organization of Securities Commissioners ("IOSCO") coordinates securities regulators. I treat this organization as just beyond the scope of this Chapter, but the reader should nevertheless be aware of its presence.

135 *See* Alan Boyle, "The Choice of a Treaty: Hard Law versus Soft Law", In Oxford Handbook of the United Nations (2019).

136 Anne-Marie Slaughter & David Zaring, "Networking Goes International: An Update", Annual Review of Law and Science 2 (2006).

137 BIS, supra note xv.

138 FSB Charter, available at this link. The FSB is technically a Swiss Law nonprofit. As noted, it is hosted by the BIS under a five-year renewable service agreement.

139 Art. 48, FATF Charter, available at this link.

140 Art. 5, Basel Charter; *see also* History of the Basel Committee, supra note xiv.

141 Art. 5, FATF Charter.

142 "BIS History - Overview," Bank for International Settlements.

143 Id.

144 "History - the BIS going global," The Bank for International Settlements.

145 *See* "Annual Report," the Bank for International Settlements.

146 "History of the Basel Committee," the Bank for International Settlements.

147 "History of the Basel Committee and Its Membership," the Bank for International Settlements, March 2001. The failure of Bankhaus Herstatt in West Germany was, specifically, the catalyst for the formation of the Basel Committee.

148 Basel Committee, Charter, Section 4, available at this link.

149 *See* Chris Brummer, "Introductory Note to the Financial Stability Board Charter," International Legal Materials 51, no. 4 (2012): 828.

150 The G20 is itself an informal organization, the convenes leaders from the world's 20 largest economies. As the Council on Foreign Relations describes it, "by gathering so many leaders together, G20 summits offer rare opportunities to develop such relationships and recast bilateral ties." James McBride, Anshu Siripurapu, and Noah Berman, "What Does the G20 Do?", Council on Foreign Relations, October 11, 2023.

151 "Report to the G20 Los Cabos Summit on Strengthening FSB Capacity, Resources and Governance," Financial Stability Board, June 12, 2012.

152 Id.

153 "Financial Action Task Force - 30 Years," Financial Action Task Force.

154 Id.

155 "IMF, Countries are Advancing Efforts to Stop Criminals from Laundering Their Trillions," International Monetary Fund.

156 "Work of the FSB," Financial Stability Board. *See* Art. 2, FSB Charter.

157 *See* "Key Attributes of Effective Resolution Regimes for Financial Institutions," Financial Stability Board, Oct. 15, 2014.

158 "An Overview of Policy Recommendations for Shadow Banking," Financial Stability Board, August 29, 2013.

159 *See* "Annual Progress Report on Meeting the Targets for Cross-Border Payments, Financial Stability Board," Oct. 9, 2023.

160 "Basel Committee Charter Article 1," the Bank for International Settlements, June 5, 2018.

161 "Basel III: International Regulatory Framework for Banks," the Bank for International Settlements.

162 "Financial Action Task Force Mandate," Financial Action Task Force, April 20, 2012.

163 "International Standards on Combatting Money Laundering, the Financing of Terrorism & Proliferation: The FATA Recommendations," Financial Action Task Force, February 2023.

164 Id.

165 "Report to the G20 Los Cabos Summit on Strengthening FSB Capacity, Resources and Governance," Financial Stability Board, June 12,2012.

166 Art. 21, FSB Charter.

167 The four committees including a Standing Committee on the Assessment of Vulnerabilities, which monitors and assesses vulnerabilities in the global financial system, and is chaired by Nellie Liang, US Assistant Secretary for Domestic Finance; the Standing Committee on Supervisory and Regulatory Cooperation (SRC), which develops policy to address key financial stability risks and coordinates issues that arise among supervisors and regulators and is Chaired by Bank of England Governor Andrew Bailey; Standing Committee on Standards Implementation (SCSI) undertakes FSB peer reviews of its members (which FSB members have committed to undergo, chaired by Bank of Japan Deputy Governor, Ryozo Himino; and the Committee on Budget and Resources (SCBR), which assesses the resource needs of the FSB Secretariat and reviews the annual and medium-term budget of the FSB and is chaired by the Chairman of Governing Board Swiss National Bank, Thomas Jordan.

168 Art. 42, FATF Charter.

169 Administrative Procedure Act, § 553.

170 "Peer Reviews," Financial Stability Board.

171 Id.

172 Id.

173 Id.

174 "Basel II Implementation," Financial Stability Board.

175 "Mutual Evaluations," Financial Action Task Force.

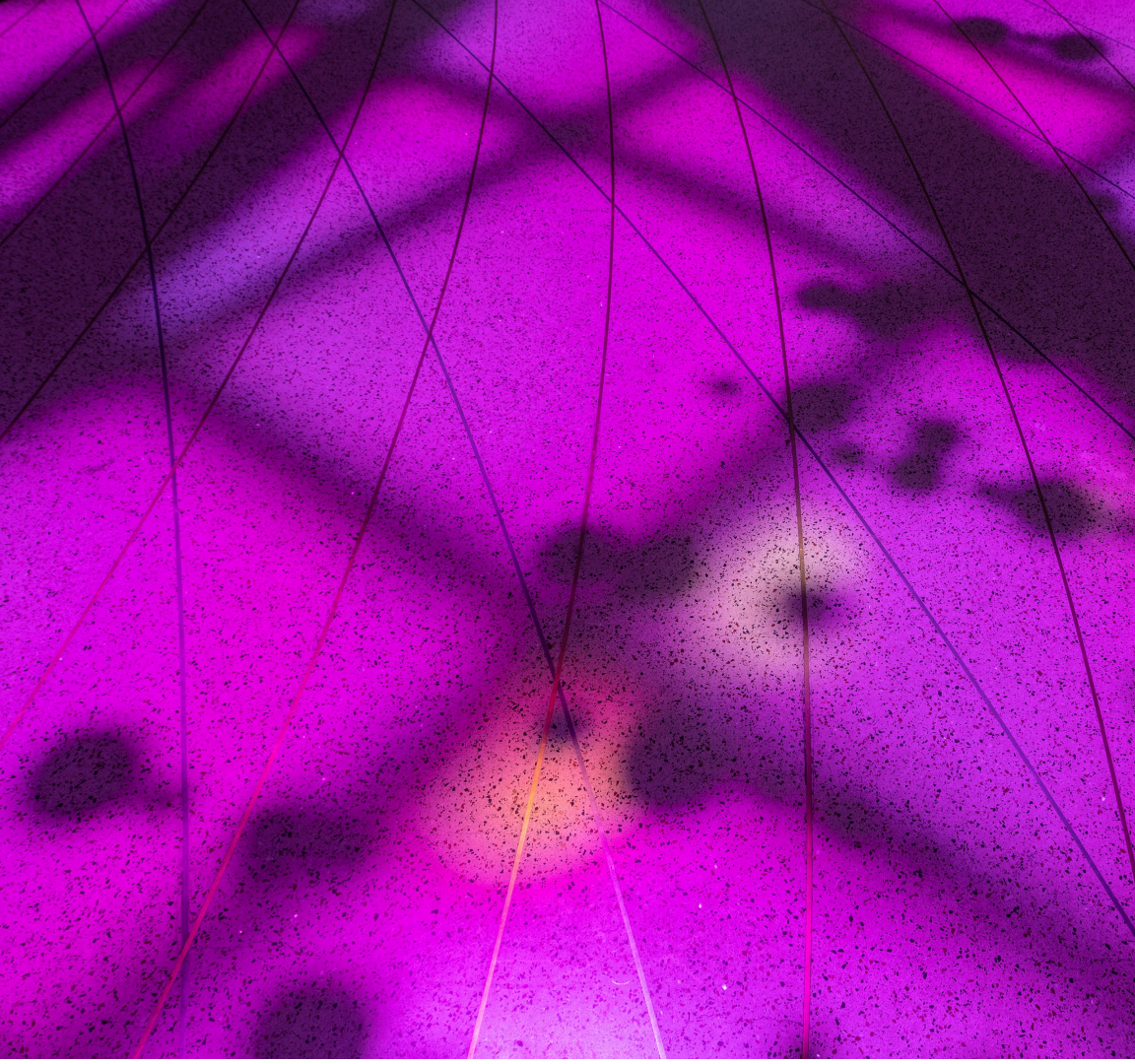
176 Id.

177 Id.

178 Id.

179 Although the FSB is formally headed beneath the G20, as it explains, "the FSB is not run by the G20—[its] membership is somewhat wider, and the FSB comes to independent policy views on issues." "Work of the FSB," Financial Stability Board.

180 *See*, e.g., Peter J. Wallison, Transparency on FSOC Designations and its Relations with the FSB (Mar. 25, 2015)," (written statement submitted to the US Senate Committee on Banking, Housing, and Urban Affairs).



Microsoft