

GOING NUCLEAR?

PRECEDENTS AND OPTIONS FOR THE TRANSNATIONAL GOVERNANCE OF AI

David Backovsky and Joanna J. Bryson

IMMEDIATELY after World War II, the United States produced almost 50 percent of the world's GDP.

America was standing at the peak of its relative economic and technological power. This technological dominance included the scientific advancements in nuclear technology that followed the Manhattan project. Although the United States did not hold onto its nuclear monopoly for long, in the post-war era it had become (and remains to this day) the principal nuclear player in international politics, creating and implementing policy that has resulted in decades of widely effective global governance of nuclear technology.

Parallels are evident with the present situation with artificial intelligence (AI), and hopefully instructive.

The debate on the validity of the comparison between nuclear and AI governance has been remarkably active over the last several months. Experts on AI ethics such as Rumman Chowdhury or Gary Marcus have been joined by industry activists such as OpenAI's CEO Sam Altman in pushing the idea of an organization like the International Atomic Energy Agency (IAEA) for AI governance. Experts in the nuclear field such as Bill Drexel, Michael Depp, Ian J. Stewart, or Matt Korda and Divyansh Kaushik, have provided counter positions, advising the AI community against modeling itself extensively after nuclear governance. These debates are important and require a concrete examination of the current state of global AI governance, looking at the actual players, as well

David Backovsky is an Associate at the Centre for International Security at the Hertie School and the host of the Berlin Security Beat podcast. You may follow him on X @BackovskyDavid. Joanna J. Bryson is Professor of Ethics and Technology at the Hertie School in Berlin, and one of the world's leading experts on AI, ethics, and collaborative cognition. You may follow her on X, mastodon, or bluesky @j2bryson, or on LinkedIn at <https://de.linkedin.com/in/bryson>. The authors would like to extend their heartfelt gratitude to Dr. Ronny Patz for his invaluable contributions during the conceptualization phase of this piece and his insightful editorial comments that greatly enhanced the overall quality of the article.



Group photo at the 2022 GPAI Summit in Tokyo

as a firmer grounding in the structural lessons of nuclear and other global technological governance.

International governance of AI may seem like the surprising topic of the moment, but it has been a concern for years, with various movements organized, legislation drafted, and international organizations formed over the past decade. Notable is the Global Partnership for AI (GPAI), which was initially formed in 2020 by G7 members France and Canada. GPAI now has 29 members. It is perhaps the international organization concerning AI with the strongest state-level support, though alternatives exist. GPAI is certainly the international organization with the greatest level of support from Western democracies currently interested in transnational impacts of AI. Following

the recent G7 meeting in Hiroshima, GPAI has been tasked with driving the Hiroshima process on AI global governance. Yet, GPAI has to date been continuously stifled by national interference, with its purpose unclear and its mandate limited. At the same time a competing and fragmented terrain of other organizations are fighting for dominance in the global space, most prominently the Organization for Economic Cooperation and Development (OECD), the International Telecommunications Union (ITU), and the United Nations Educational, Scientific, and Cultural Organization (UNESCO, which the United States has recently rejoined) each providing a different aspect of expertise and capacity. Neither GPAI nor these other candidates are prepared to take on the role necessary for AI global or even transnational, regional governance.

In this article, we examine the current state of AI transnational governance. We explore the remit and challenges of GPAI. We then weigh in on the nuclear-AI debate, examining which of the many lessons available from nuclear governance are instructive, and where this parallel stops. We provide counterarguments to the recent critiques of nuclear governance as a model for AI. Our core argument is that the United States in particular does not seem to be deploying essential lessons from its own past and present leadership on the governance of emerging technologies. U.S. strategic leadership on nuclear governance, however fraught that regime has been, should prove instructive on the need for Washington to lead on—but not monopolize—global AI policy and governance as well. Other players have already shown competence and willingness to take national and regional initiative. Ensuring a coordinated and transparent approach will be essential to a digital future that promotes widespread human thriving.

THE GLOBAL PARTNERSHIP FOR AI

Digital technologies (DTs) have long been touted as great potential equalizers of economic development, and it is true that the world overall is more equal than it was in 1945. In

particular, the U.S. share of world GDP has now shrunk from more than half to slightly less than a quarter. However, although fears abound about the United States losing its technological and economic dominance, as Google’s former CEO Eric Schmidt describes in a February 2023 essay for *Foreign Affairs*, in the field of AI America possesses a strength strongly comparable to its early dominance in nuclear technology—in development, scale, and computing power. The current terrain of global governance of AI is deeply fragmented. Different organizations compete for dominance and legitimacy. Industry founders bringing in leading NGOs and universities created the Partnership for AI in 2016. The ITU has been working on gaining a lead in the technical governance of AI and has established cooperations with UN organizations, such as the World Health Organization (WHO). Supranational organizations including the European Union, OECD, the United Nations Development Program, and UNESCO have invested considerable resources into governance instruments and are jockeying for position.

Here we focus on one entity in particular, GPAI. This is partly because of GPAI’s special position within

U.S. strategic leadership on nuclear governance, however fraught that regime has been, should prove instructive on the need for Washington to lead on global AI policy and governance as well.

The current terrain of global governance of AI is deeply fragmented.

Nevertheless, the GPAI terms of reference include a statement demanded by the Trump administration, that the Partnership could engage in no ‘normative’ activity:

2.7, page 3 “GPAI focuses on the application of AI and the implementation of the principles set out in Annex A, rather than on developing high-level AI norms or policy. GPAI will not work on issues of national defence.”

The original understanding of this text was that the GPAI would not be granted normative powers such as those wielded by the EU, where the union is able to agree on binding policy recommendations that then must be crafted into laws of each member state. However, the GPAI steering committee has at least for its first three years chosen to interpret this language as meaning that the governance of AI, taken broadly, was not an admissible topic for expert

powerful nations. The May 2023 communiqué of the G7 Summit at Hiroshima calls to “advance international discussions on inclusive artificial intelligence (AI) governance and interoperability to achieve our common vision and goal of trustworthy AI, in line with our shared democratic values” and tasks GPAI with establishing the “Hiroshima AI process.” A second consideration for our focus is that one of the authors (Bryson) was nominated as one of the first cohort of GPAI ‘experts’—a role described further below. She was nominated by Germany, a country in which she is resident but not either of the ones of which she is a citizen (the United Kingdom and the United States, both of which are also GPAI members). As such, we have had considerable opportunity to discuss GPAI with a wide range of important stakeholders and in a variety of contexts (though many were discussed under the Chatham House Rule).

GPAI was originally proposed in 2017 as a Canadian and French initiative. The proposal was received with skepticism, particularly by the Trump administration. Major fears were expressed regarding regulation’s tendency to hamper innovation (though in fact as has been long and well established, monopoly and size stifle innovation, and regulation is an essential antidote). The initiative was apparently originally expected to be an AI version of the Intergovernmental Panel on Climate

study. This was an astonishing interpretation to external observers, for example at the UN, who had understood governance to be the GPAI’s primary intended role.

The primary role communicated initially to the experts, at least, was that GPAI should produce attractive public goods in the area of AI that would encourage more nations to want to join the family of liberal democracies. Language about liberal democracy was rapidly depreciated though, as Singapore took a leading role in

As a technology, AI’s “nature” is far less separable from its governance context than that of biology, weather, or ecosystems.

GPAI, providing significant value from their own expertise in digital governance, and also with efforts to reach out to nations such as Türkiye and Egypt. Besides the United States, the other keenly sought and hard-won initial member was India. The initial membership thus looked much like some proposals for G7 expansion. Interestingly, the present, 29-state membership of GPAI overlaps with some of the expansion members of the Shanghai Cooperation Organization (SCO), an organization led by Russia and China, focusing since 1996 on military, and since 2006 also on Internet security. Although the SCO initially included only in addition, Kazakhstan, Kyrgyzstan, and Tajikistan, it has expanded to include both India and Pakistan as full members. Besides

India, other GPAI members and observers also hold or have applied for “dialogue partnership” status with SCO, including Türkiye and Israel.

Shortly after GPAI’s formal public launch on June 15th, 2020, *Politico* published an article by Janosch Delcker with a very different view of GPAI’s remit. Entitled “Wary of China, the West closes ranks to set rules for artificial intelligence,” the article claims the primary concern of GPAI would be addressing China’s AI ascendance. In the words of then U.S.

Deputy Chief Technology Officer Lynn Parker, GPAI should be “a good counter to China.” Were addressing concerns of China’s AI policy really the primary concern, perhaps a G20-style organization would be a better venue than a G7 one. In terms of AI competence, China certainly deserves to be treated as a potential partner; it certainly impacts intelligent technology globally.

At least a few GPAI experts were skeptical of *Politico*’s Chinese interpretation of GPAI’s unstated remit, believing instead that corporations were the “other” against which GPAI had been organized to gain competence in negotiating. Governments were seen as relatively technologically ignorant, and were not being taken seriously by the corporations

producing the main impacts with AI. This suggestion relates to yet another possible motivating concern for the partnership, one we have not seen mentioned elsewhere: the question of why the United States itself stopped applying its own laws on governing the kind of market dominance we see in many AI sectors. For example, 80 percent of search goes to a single company, Google. Twitter (now called X) is (among other things) a dominant mechanism of political communication—some governments used it as their sole, trusted system for disseminating COVID-19 information—a trust that was apparently deserved, at least up until its recent private purchase. Apple holds a share of ca. 27 percent of the smartphone market worldwide, and around 50 percent in the G7. When after World War II, the allies (particularly the U.S. and the UK) forced Germany and Japan to adopt competition law similar to their own, the explanation was that if corporations are allowed to become too large, they either take over the government or are taken over by the government, either of which results in autocracy. The Trump administration has often been described as having had autocratic leanings, and indeed apparently considered whether the U.S. government should have a closer relationship with its digital sector, “like some other countries.”

In terms of AI competence, China certainly deserves to be treated as a potential partner; it certainly impacts intelligent technology globally.

Regardless of the original intent for GPAI, its first three years have proven deeply frustrating for many experts, though some have thrived. Most organizations in our experience bring in experts when they have known questions or policy domains they need to address. In our experience, the UN, ICRC, Chatham House, WEF, and even software companies like Google or Meta, bring together experts for a short period, ply them with questions (and sometimes good food), allowing the experts to raise issues and hash out differences of opinion between themselves, under observation by the organizers. Great intellectual progress can sometimes be made surprisingly rapidly if the right assortment of varied expertise is assembled and well facilitated. The organization will itself then take ownership, and indeed often authorship (as experts are ordinarily shielded by the Chatham House Rule), of generating any outcomes. This strategy allows organizations to produce policy in light of expertise, including importantly the organization’s own. The organizers, rather than the experts, may best know their own political reality—what can be achieved given that organization’s available resources and priorities.

In GPAI, in contrast, experts in most countries were isolated from their nominating authorities. Indeed, the GPAI

expert system has now been altered to favor “self-nominating” experts with even less legitimacy for their actions, or information on concerns, derived from the actual GPAI members. At the first two annual meetings, experts found themselves relegated as overqualified passive members of an audience, entirely cut off from the ‘partners’ they were supposed to be consulting to—a situation improved by the third chair, Japan, in the 2022 Tokyo meeting. Outside the annual meeting, throughout most of the year, the experts were instructed to select their own problems to work on for as much time as they were willing to volunteer, with desired outputs initially unclear but eventually reduced to producing reports as deliverables. Experts were first asked to produce proposals for these projects, which the steering committee then selected entirely opaquely—even chairs of working groups had relatively little discretion to allocate resources. In some cases, a very small number of actors received disproportionately large amounts of meager available resources, including speaking slots in the first two annual meetings.

Unquestionably, some excellent work has been achieved by experts in this GPAI context, but the question remains whether this is the best way to derive or deploy AI expertise, let alone whether it is a process that might be used to coordinate global or transnational AI governance. An inordinate amount of time

was spent by the actual GPAI members (not experts) at the Tokyo meeting in an unsuccessful attempt to convince one of their members to accept a membership application of another important nascent democracy. Digital governance is of too much immediate import to bog down rare meetings in such petty acts of diplomacy largely irrelevant to the task at hand.

THE ATOMIC DEBATE

The discussion regarding the extent to which nuclear governance should serve as a guide for the development of the AI regime has been generating attention and discussion in public discourse at least since *Wired* published Rumman Chowdhury’s April 6th, 2023 article with a compelling title: “AI Desperately Needs Global Oversight.” We believe that to date this debate has missed the larger picture, regarding where the nuclear governance regime and the IAEA prove instructive and where the comparisons stop, as well as important larger structural lessons of nuclear governance. Many of the critics of nuclear as a model for AI make arguments that we believe in fact prove how valuable such an approach could be.

The general counterarguments run as follows: nuclear governance has been and remains a fraught regime with many successes, but also significant setbacks and issues. The beginnings of the global governance of nuclear technology were,

for example, laden with failed attempts. The UN’s Atomic Energy Commission, established by the very first resolution of the UN General Assembly in 1946, ended in failure as the Soviet Union vetoed the so-called Baruch plan to impose limits on the development of nuclear weapons and technology. The IAEA was established only 11 years after the nuclear explosions at Hiroshima and Nagasaki. It has also been argued that nuclear governance has been built only in response to crises, leading us to cycles of learning lessons only from our mistakes, and that this approach could prove disastrous for AI governance. Moreover, it has also been argued that rather than cooperative global governance, it has been American strategy to contain nuclear technology that has proved the strongest force in mitigating nuclear risks.

All this history is predominantly correct, and certainly we cannot wait 12 years to establish an AI control regime. It would be wrong to say we have learnt solely from our errors, but also wrong to think that we can ever ensure flawless behavior, or that learning after a mistake is a poor outcome. Besides claims rooted in history, critiques have also been focused on fundamental differences between nuclear and AI technologies, which are also incontrovertible, but perhaps less salient to technology governance than they might seem. Some have put forth the argument that while the paths leading to risk in AI remain

uncertain, the pathways in nuclear systems, such as nuclear proliferation or nuclear war, are considerably more evident. Others suggest that AI systems are much more prone to proliferation than nuclear systems, or that AI systems, being digital, are more ephemeral.

Although these counterarguments sound compelling, they overlook some equally undeniable and relevant facts. Overall, the IAEA and the non-proliferation regime have been one of the most successful stories of international governance of emerging technologies the world has seen. The IAEA is widely considered one of the most efficient and effective international organizations in existence, one that would have to be invented if it didn’t already exist. While some have argued that nuclear technology is restricted to energy and weapons technology, the IAEA in fact deals with an incredibly broad range of uses of nuclear tech: from energy and medical uses, to agriculture, law enforcement, and more. It is capable of conducting audits of public and private installations in a field that requires high levels of expertise in nuclear physics. Its laboratories, such as in Seibersdorf in Austria, are world-class nuclear installations.

We must also remember that the IAEA not only provides safeguards, but extensive technical coordination capacities to states and their industries. Certainly, we can imagine such technical coordination

capacities generated by a comparable international AI organization helping the responsible implementation of AI transnationally. Importantly, the IAEA also integrates its political and technical functions into a single capability. This is something that does not yet exist in the world of AI global governance, and indeed represents something against which the United States has been defending, at least in GPAI. This is, in our estimation, an error.

Artificial intelligence is indeed primarily digital, but digital technology is physical. All computation, digital or biological, is a physical process of changing information, which information (data) must also be physically manifest. Computation requires time, space, and energy. Data also requires space and energy for its storage. The vast infrastructure underlying our AI capacities requires both physical and cyber defense, indeed quite similar to what nuclear power plants require. Presently, much artificial computation is done on advanced semiconductor chips, and in fact U.S. export controls on advanced chip technologies already resemble technology controls on certain nuclear equipment.

A nuanced discussion of technology governance requires more debate regarding what makes the IAEA so effective as

an organization. Research shows that the IAEA benefits from a centralized structure with a politically insulated secretariat that has sufficient autonomy to remain an honest and legitimate broker. A notable demonstration of the IAEA's political independence occurred in 2003, when the then Director General, Mohammad El-Baradei, stood against the United States, affirming to the UN Security Council

The United States should want to use its dominant position in the AI commercial sector to become a leader and drive the establishment of a legitimate governance regime.

that no weapons of mass destruction were found in Iraq by IAEA inspectors. Yet it is also incontrovertible that the United States dominates the securing of nuclear power, waste, and weapons systems, through the investments of its Department of Energy.

It is this mixture of autonomy, legitimacy, and political-technical capacity that we believe should be the core takeaways from nuclear governance for the AI regime. We argue that these structural lessons will prove much more important than unnecessarily scrupulous comparisons between less relevant differences in nuclear and AI safeguards.

STRATEGIC AND READY

Whether by design or accident, Eisenhower's decision in 1953 to push the Atoms for Peace initiative helped lay the groundwork for effective nuclear governance around the world.

Although the debate among nuclear scholars remains contentious, the non-proliferation regime is arguably one of the strategic successes of the United States. It was the combination of American strategic leadership and a legitimate global institution, which permitted this regime to work. This is the key lesson we might all take away, but particularly the United States. It is essential that the United States looks back towards its own success and applies these lessons to the international environment now. It is insufficient to debate the domestic regulatory environment alone, the United States should want to use its dominant position in the AI commercial sector to become a leader and drive the establishment of a legitimate governance regime. Given its present weaknesses in regulation, it might though be wise to choose to respect and partner with those moving forward faster—certainly the EU, but possibly also China.

Chinese technological prowess also played a key role in the development of the present nuclear non-proliferation regime, a story that offers further lessons for AI governance. The 1964 Chinese nuclear test at Lop Nur represented the culmination in a worrying increase of nuclear proliferation in the decades after World War II. The mid-1960s saw the United States, along with the rest

of the international community act. Collaborative efforts were undertaken resulting in the formulation of the Non-Proliferation Treaty (NPT), signed in 1968. The sweeping transformation that the IAEA underwent during this period, as captured by Elisabeth Röhrlich in her 2022 book entitled *Inspectors for Peace: A History of the International Atomic Energy Agency*, was noteworthy for its significant enhancement of nuclear safeguards. It's crucial to understand that the IAEA's eventual triumph as an institution overseeing safeguards indeed largely hinged on the time it had to develop

its expertise and strengthen its legitimacy. It was this grace period that allowed the IAEA to mature into its pivotal role. When the NPT came into force in 1970, the IAEA was ready to safeguard it. Although we can hopefully accelerate this process using the lessons learned from such prior successes, the need for a grace period for AI regulatory organizations to mature unquestionably strengthens the urgency to promptly establish a global AI agency now.

We also of course agree with the critics of the nuclear governance metaphor that the nuclear governance regime was dependent on American strategic leadership. At least in the case of nuclear regulation though, neither U.S. leadership nor the IAEA and the NPT could

AI governance will require strategic leadership, including but perhaps not limited to that by the United States.

have existed without each other. AI governance will also require strategic leadership, including but perhaps not limited to that by the United States.

CAN AMERICA LEAD ON AI?

The current global governance regime for AI is deeply dysfunctional. GPAI, OECD, UNESCO and ITU form a set of competing actors that do not presently provide us with the necessary legitimacy and centralization that would best serve global governance of AI. Our experience with GPAI shows that its institutional design as it stands is not sufficient for the task at hand.

But we do have a history of technological governance, in which we can look for patterns and construct rhymes, even if the pace of technological expression is accelerating. We know the basic structure of what an organization that can govern an emerging technology can look like. Regardless of the fact that nuclear and AI safeguards will never prove to be completely comparable, many core lessons from the IAEA are straightforwardly applicable. We need a centralized agency, with political and technical capacity, with internal expertise and the right balance

Regardless of the fact that nuclear and AI safeguards will never prove to be completely comparable, many core lessons from the IAEA are straightforwardly applicable. We need a centralized agency, with political and technical capacity, with internal expertise and the right balance between accountability and political autonomy.

between accountability and political autonomy. The IAEA also teaches us that the sooner we create such an organization, the better. When the international community finally comes together to create an international regime or treaty

to govern AI, we will need an agency capable of implementing it.

Some say that the time for creating a global cooperative regime is not right. Let us remember that the IAEA was created at the outset of the Cold War and the non-proliferation regime at its absolute height. Further we can add another story of the founding of global governance, one that indeed returns to that originally contemplated

model for the GPAI, the IPCC.

Although it is sometimes forgotten, the IPCC was founded by a fairly anti-climate Reagan administration in 1988. In fact, in 1985, Reagan had wholly rejected the idea of an IPCC-like organization for climate research. However, through the work of American and international bureaucrats, the Reagan administration was convinced on relatively short order that such an organization would have value. As described by the international

organizations scholar Tana Johnson in her 2014 book *Organizational Progeny*, the IPCC was an unlikely success. The trick was that the IPCC was built on top of a mostly dysfunctional framework of a political and technical agency—the United Nations Environment Programme and World Meteorological Organization, respectively. Something new of global import was built on top of two things that did not work, despite strong initial American resistance.

Drawing upon such valuable insights, we possess a broad blueprint for managing emerging technologies. For one thing, it becomes evident that the existing framework, where GPAI assumes responsibility for the new Hiroshima AI initiative, is fundamentally flawed under GPAI's present limits. Therefore, it is imperative to either empower GPAI as the centralized and effective agency required such as the IAEA, or construct a new one, as per the IPCC, involving key stakeholders such as UNESCO, GPAI, and the ITU. First indications of such solutions are already appearing with the return of the United States to UNESCO and the UNSC's first meeting on generative AI in July 2023.

Though AI is often presented as a natural entity (see previous mention of an alleged science of “machine behavior”), in fact it is a set of engineering techniques with diverse economic and security potential. Nations which wish to lead, including the United States, should

carefully examine the successes achieved not least by the Americans in strategically and yet cooperatively driving nuclear governance and the non-proliferation regime, and apply these lessons to the current failings of GPAI and other candidate agencies. Since the beginning of its post-colonial history, American leadership has resulted in robust international institutions that often maintain their legitimacy for decades. In the last century, whenever the United States decided to strategically support global governance, this served as a catalyst, enabling the international community to form regimes that have been largely effective in governing emerging and potentially dangerous technologies. Now as in the early nuclear age, the United States has the opportunity to lead in the global governance of a technology it presently commercially dominates. Yet presently, America seems to be desperately failing to learn from this history in its approach to AI. We hope this article will encourage the U.S., its allies, and indeed governments globally to recognize the importance of these precedents, and to come together even more swiftly on a regulatory strategy. Like nuclear power, AI has the potential to not only cause problems, but vastly more so to help us solve them. Investing in an intense push to consolidate effective mechanisms of digital governance and digital cooperation may indeed lead to more rapid, even conclusive progress on other globally shared problems. ●