UNIVERSITEIT VAN AMSTERDAM

regulAite

Global, Selective, or Both? The Case for Differentiated Cooperation in AI Governance

Nora von Ingersleben-Seip

Daniel Mügge

RegulAite

working paper

03/2025

Abstract

Current debates about international cooperation in AI governance remain both simplistic and muddled, pitting global forms of collaboration against selective alliances among "like-minded countries". We propose a more nuanced and systematic approach to cooperative AI governance based on three considerations. First, different kinds of governance issues lend themselves to different kinds of cooperation. Second, not all AI is created equally: different kinds of AI raise different challenges, thus requiring varied forms of cooperation. Third, the same is true for the development and deployment phases of AI systems. Integrating these three factors, we argue why some challenges can only be addressed through global cooperation, while for others selective cooperation is an equally effective, or even more effective, solution. In consequence, governments should not insist on either global or selective cooperation. Instead, they should opt for the scope of collaboration that is most effective for solving any particular governance issue at hand.

How to cite this paper

von Ingersleben-Seip, Nora and Müggе, Daniel (2025) *Global, Selective, or Both? The Case for Differentiated Cooperation in AI Governance*. RegulAite working paper 03/2025, Amsterdam.

Author information

Nora von Ingersleben-Seip is a postdoctoral researcher at the University of Amsterdam. She works on the RegulAite project, investigating how and when the European Union can collaborate with China in AI rule design and enforcement.

Daniel Müggе is Professor of Political Arithmetic at the University of Amsterdam. He researches the Politics of AI governance in the EU and beyond. He can be reached at d.k.muegge@uva.nl.

Author email addresses

Nora von Ingersleben-Seip: n.a.voningerslebenseip@uva.nl.

Daniel Mügge: d.k.muegge@uva.nl.

Corresponding Author

Nora von Ingersleben-Seip, University of Amsterdam (n.a.voningerslebenseip@uva.nl).

Statement regarding Conflicts of Interest

The manuscript is original, has not been published, and is not under consideration elsewhere. I confirm that my co-author has approved the manuscript and agrees to its submission. We have no conflicts of interest to disclose.

RegulAite is a Vici research project funded by the Dutch Research Council (NWO, grant number VI.C.211.032), hosted by the University of Amsterdam. RegulAite is led by Prof. Dr. Daniel Mügge and investigates EU AI governance in the context of global geopolitical and economic competition. More information can be found on the RegulAite website.

Introduction

The opportunities and challenges created by AI have inspired governments to intervene in its development and deployment. Beyond unilateral efforts—in which we include EU-level initiatives—jurisdictions have launched international cooperation in AI governance, for example through the Organization for Economic Cooperation and Development (OECD) or the United Nations (UN).

Such cooperation remains contested. Some policy and academic analyses claim that effective AI governance requires global cooperation, *tout court.* For example, the UN High-Level Advisory Body on Artificial Intelligence in its 2023 interim report called for an encompassing global governance framework on AI.[1] In principle, such initiatives could be combined with selective cooperation at the regional level. But some argue that a global "focal organization" for AI governance is necessary, subsuming many of the already existing, smaller-scale AI governance initiatives.[2] Thus, in its 2023 interim report, the UN AI Advisory Body declared categorically that *"AI governance must be universal"* and that "new horizontal coordination and supervisory functions are required and they should be entrusted to a new organizational structure".[3]

Such plans to make the UN Secretariat the superstructure for AI governance quickly elicited a backlash, and Western think tanks and research institutes have warned that cooperation with authoritarian countries—especially China—on AI governance rules would be undesirable.[4] Experts point to China's behavior in the International Telecommunication Union (ITU) and its perceived hypocrisy when it comes to the development of lethal autonomous weapons (LAWS) as reasons why China cannot be a trusted partner in AI governance.[5] Both US and European policymakers frequently cite value clashes as reasons why democracies cannot collaborate

with non-democracies on AI rules.[6] Taken together, the discussion pits selective against global cooperation as mutually exclusive approaches.

To our mind, this debate is too simplistic for three reasons. First, different kinds of governance issues lend themselves to different kinds of cooperation. For example, global public goods problems are best dealt with by international organizations with (nearly) global membership, while military alliance building by its very nature requires selective cooperation within smaller organizations. Second, the AI moniker subsumes such diverse digital systems that different cooperation rationales for them—stemming from the different governance issues they pose—are only to be expected. Large language models (LLMs), for example, raise very different challenges than AI-powered home appliances. Since the issues that need governing in the case of AI vary enormously, so do the rationales for collaborating and for embracing one or the other scope for collaboration.[7] And third, governance challenges—and hence incentives for cooperation and the most effective ways to do so—vary significantly between the *development* phase of an AI system and its subsequent *deployment*. States may therefore need to prioritize one kind of cooperation in the development phase and another in the deployment phase. Against that background, we argue, cooperation in AI governance needs to be approached differently: when and where is global cooperation necessary to tackle governance challenges, and when is selective cooperation an equally viable (or even the better) solution?

To answer this question, we use a deductive approach, marrying theories of cooperation and typologies of public goods to an understanding of the diverse challenges AI can pose. We use this approach for two reasons: first, much of the debate is about the potential and desirability of cooperation, given that cooperative AI governance is still nascent and the empirical track record therefore thin. Second, the deductive approach allows us to isolate the dynamics we are interested in here—the problem-solving potential of different forms of cooperation—from

other factors such as rising techno-nationalism, which confound observable patterns of cooperation in AI governance.[8]

In a first step, we disentangle potential motivations for cooperative AI governance. We then build a framework to evaluate the up- and downsides of selective and global cooperation with reference to each of those motives. That allows us to see when global cooperation is necessary—and when selective cooperation is an equally or even more effective solution. While private actors can—and do—play an important role in AI governance, we concentrate on state actions, as only states and multilateral bodies are able to craft binding national or international laws.

In total, we analyze seven motivations for interstate cooperation on AI governance, drawing on theories in International Relations (IR) and International Political Economy (IPE): building or bolstering geopolitical alliances; avoiding global arms races/races to the bottom; preventing the proliferation of harmful AI; realizing benefits from trade; sharing knowledge and best practices; promoting norms; managing externalities of harmful AI and collective action problems.

When governments need to avoid races to the bottom, truly global collaboration is required—certainly among all major powers.[9] In contrast, something like mutual market access can be pursued bilaterally, even if global collaboration might be preferable. Yet other motives, such as alliance-building, actually require selective cooperation. Because many of these issues can be tackled independently from each other, countries need not choose between the two approaches to AI governance. Instead, a *differentiated* approach to cooperation makes much more sense. Thanks to the diverse governance challenges raised by different kinds of AI in different AI lifecycle phases, working with a select number of partners in some domains, with a larger circle in others, and pursuing global solutions in yet others is the best strategy.

Figuring out where selective cooperation in AI governance is most effective and where global cooperation is the better (or maybe the only) option is important for two reasons. From an analytical perspective, it helps us understand AI governance challenges better, which of them are most salient in which phase of the AI lifecycle, and with which governance mode governments can best address them. From a practical perspective, our arguments help policymakers decide when to pursue global AI cooperation, and when to aim for selective cooperation instead.

This working paperarticle proceeds as follows: we first delimit the scope of our inquiry and sketch the dominant debate about AI governance. Based on the IR and IPE literature, we then discuss seven motivations for cooperation in AI and spell out their implications for AI governance challenges. The subsequent section zooms in on one of those categories—externalities from cross-border AI harms—and shows how we need to differentiate between the development and deployment phases of the AI life cycle and different forms of AI to pinpoint which kinds of cross-border governance may be needed. The final section combines these elements in a heuristic for understanding AI governance cooperation and spells out how our findings argue in favor of a variable approach to cooperation, rather than one that opts for either the global or the selective variant.

Delimiting AI Governance

AI is an amorphous and misleading concept. In practice, it has most closely been associated with machine learning, but it also includes various symbolic approaches to building systems that attempt to emulate human thinking.[10] We use the label despite this blurriness because it dominates public and policy debates. “AI” structures countless working groups, standard-setting bodies and rule-setting initiatives at both national and international levels.[11] AI is real

as a social construction that, as such, leaves a footprint in the world, including on how the technologies that fall under AI are governed.[12] At the same time, the term injects unhelpful and un-reflected slants into policy debates. The sundry technologies lumped together under the AI heading invite blanket statements about the desirability of "cooperation "in AI"—ignoring that automated target identification for military drones is a completely different matter than smart thermostats to keep the kitchen warm. A key part of our mission is to disentangle the AI knot as relevant to governance cooperation, an issue to which we return in the penultimate section of this working paper.

Our focus makes us pragmatic in how we delimit our field of inquiry. In principle, both public and private actors feature in AI governance. But because we care about interstate cooperation, AI governance here encompasses public actor interventions to shape the development and deployment of AI technologies. Cooperation on AI governance, then, means international coordination of these public interventions. That could take many forms: agreements about facilitating knowledge exchange, attempts to align regulation to smoothen mutual market access, direct sharing of military AI with allies, or non-binding dialogues among government officials about the limits to responsible AI use.

Cooperation ranges from two jurisdictions working together to something covering the whole globe. Lest matters get unduly complicated, we distinguish two scopes for cooperation here: (more or less) global cooperative AI governance, which certainly includes all the major AI powers (and crucially China and the United States); and limited cooperative AI governance ("selective cooperation"), which includes at least two jurisdictions but misses at least one central AI power.

We observe both global and selective cooperation initiatives in the wild.[13] With a global scope, we find the G20 countries (which include China and Russia) and the UN General Assembly,

which has promoted ethical, safe, secure, and trustworthy AI among its 193 member states through its resolution on AI for the global good.

On the selective cooperation side, the OECD's 36 member countries plus Argentina, Brazil, Colombia, Costa Rica, Peru, and Romania signed the OECD AI principles in 2019.[14] Other limited initiatives include NATO's 2021 AI strategy, the G7's 2023 Hiroshima Principles, the Council of Europe's Framework Convention on Artificial Intelligence and the Ministerial Declaration of the Global Partnership on AI from 2024, with 29 signatories.[15]

Also in that category we find transatlantic cooperation through the EU-US Trade and Technology Council (TTC), founded in 2021 to "advance Transatlantic cooperation and democratic approaches to trade, technology, and security".[16] Already before Donald Trump returned to the White House, the TTC lost momentum, and it has done yet more so since early 2025.[17] Nevertheless, the transatlantic axis remains a central potential link in global AI governance, considering how the United States and the EU have stood at the center of multilateral technology governance initiatives.

Debating Cooperation in AI Governance

Most advocates of cooperation in AI governance fall into one of two camps: some see global cooperation as the only way forward; others champion strategic alliances, instead. In the first camp, the UN AI Advisory Body in its 2023 interim report declared categorically that "AI governance must be universal" and that "new horizontal coordination and supervisory functions are required and they should be entrusted to a new organizational structure".[18] Chinese foreign minister Wang Yi, too, called for an "international AI governance institution [to] be set up under the UN framework". Cooperation in a smaller circle of countries ("small yard, high fence," as he described it) would "result in mistakes with historic consequences".[19]

Also in academia, we find champions of global AI cooperation. Pekka Ala-Pietilä and Nathalie Smuha, for example, find such cooperation necessary to protect citizens, support socially beneficial innovation, and safeguard market competition.[20]

Other scholars and policymakers promote selective cooperation, centered on alliance-building with "like-minded" countries—a grouping that typically involves more or less democratic US allies, shutting out others who are thought to have different goals and values.[21] (We remain agnostic for now here whether those concerns are genuine or only facades to hide hard-nosed geopolitics or economic competition.)[22] In 2021, the Biden-Harris administration announced its intention to bolster tech cooperation among democracies against authoritarian regimes.[23] Since then, democracies have cooperated on AI standards mostly within the Global Partnership on Artificial Intelligence (GPAI) and the OECD (efforts that effectively merged in 2024), rather than, for example, the UN. GPAI explicitly restricts membership to countries that endorse "the shared values reflected in the OECD Recommendation on AI or, for applicants that are not adherents to the OECD Recommendation, the principles on Artificial Intelligence in appendix A to the GPAI Terms of Reference".[24] Appendix A to the GPAI Terms of Reference outlines values and principles for the responsible development and use of AI, which include respect for human rights, inclusion, and a commitment to democratic principles and international cooperation. China, so the implication, is not welcome at the table.

And then, there are those who warn that a brewing AI arms race would make global cooperation in the field outright dangerous. This alleged arms race incentivizes countries to develop military AI as quickly as possible, and anything that slows it down—including selective cooperation with like-minded countries—only helps the adversary. We find such arguments, for example, from former U.S. Secretary of Defense Mark Esper and former Pentagon Chief Software Officer Nicolas Chaillan.[25] More recently, the second Trump administration has also embraced such arguments again.

But to what degree do these encompassing positions make sense, given the diversity of cooperation incentives, the different challenges raised by different AI lifecycle phases, and the many kinds of AI out there?

What Motivates States to Cooperate on AI Governance

Scholars of IR and IPE have put forward different theories of international cooperation.[26] Viewed pragmatically, these approaches are neither entirely right or wrong, but capture different dynamics that are more or less pronounced in different policy fields. Constructivist approaches have particular purchase on real-world dynamics in fields in which norms and uncertainty play a big role; realism has more to say where security dilemmas loom large. Neoliberal institutionalism is particularly useful when repeated positive interactions have led to high trust among countries, enabling the creation of public goods; mercantilist theories highlight cooperation patterns that emerge when national economic interests are thought to collide. To understand the varied motivations for states to cooperate on AI governance, we first explore to what kind of governance challenges these theories typically apply, and what they suggest about the cooperation dynamics we should expect and that would be most apposite.[27]

Neorealist IR scholarship holds that, because of the anarchic nature of the international system, states are motivated to cooperate only when it helps them preserve or gain relative power.[28] Thus, cooperation is driven by a desire to constrain rivals to prevent them from gaining a competitive edge. For example, the United States might support a global treaty banning autonomous weapons if doing so will prevent geopolitical rivals such as China from gaining a technological edge in their development. Since Neorealism also assumes that, due to the absence of a central authority above states, international institutions are weak and trust among states is low, there would have to be strong monitoring and enforcement mechanisms

ensuring that countries stick to their commitments. In this zero-sum logic, cooperation is driven by narrowly defined self-interest and characterized by temporary alliances rather than deep, institutionalized cooperation.

Neoliberal institutionalism sees states cooperate because doing so helps them to manage the complex interdependence that arises from the transboundary nature of AI. For example, states might cooperate on rules for international data flows, AI-driven cyber threats, or global supply chain issues. Neoliberal institutionalism has particular purchase on issues that for which cooperation benefits everyone.[29] It assumes that repeated interactions and the establishment of institutions that help lower transaction costs, set expectations, and monitor compliance increase the chances of successful cooperation among states. Once such institutions are in place, states can work together on creating common standards, fostering data sharing, and preventing harmful uses of AI.

From a constructivist perspective, states cooperate because of shared ideas, identities, and norms.[30] Thus, like-minded countries might work on joint AI governance frameworks that reflect values they hold in common; they might for example seek to promote "human-centric AI" in the name of freedom and dignity. Constructivists assume that states' interests are not fixed but shaped by social interaction. Therefore, engaging in dialogue with countries such as China or Russia might gradually socialize such countries into liberal-democratic norms.

Finally, mercantilist IPE sees states cooperate to secure strategic industries, control technology value chains, or protect national champions.[31] Cooperation becomes a tool of industrial policy, not just interdependence management. For example, the EU and the United States might align on AI standards to obstruct Chinese tech firms, preserve market dominance, and control data flows. AI governance then serves the goal of strategic decoupling or reshaping

global capitalism. Mercantilist IPE provides a more structural and power-sensitive view of AI cooperation that looks at economic dominance, industrial rivalry, or inequality.

Taken together, this literature points to varied motivations for cooperation: building or bolstering geopolitical alliances; avoiding global arms races/races to the bottom; preventing the proliferation of harmful AI; realizing benefits from trade; sharing knowledge and best practices; promoting norms; managing externalities of harmful AI and collective action problems. But to what degree are either selective or global forms of cooperation effective to tackle these, meaning that they have a high chance of countries actually attaining their intended goals? To answer that question, we explore these seven goals in a bit more detail.

Building or Bolstering Geopolitical Alliances

AI upsets existing international power constellations, and by reshuffling offensive and defensive capabilities, it also affects the probability of war.[32] Moreover, alliances like NATO reassess their strategies: how should its members integrate AI into military operations and counter new digital threats?[33] As the digital and material dimensions of warfare increasingly intermesh, allied countries face incentives to share information, mutually adapt systems, and trade in the relevant components to exploit economies of scale.

Major AI powers (notably the NATO countries, China, and Russia) can forge alliances with other strategically positioned countries; countries that fall outside that category simply may have to bandwagon as they scramble to update their military capabilities. If digital military systems are increasingly integrated and connected, third countries will find themselves increasingly locked into de-facto ecosystems of compatible components produced by the major AI powers.

When it comes to bolstering geopolitical alliances, only selective cooperation makes sense. Note, however, that that applies only to those kinds of AI that have plausible geostrategic

implications, such as military use, not to those without such implications. Moreover, the argument against broad and potentially global cooperation also does not hold when it concerns kinds of AI that have already been developed by geopolitical competitors, as well. The parallel with semiconductors is instructive: the United States has banned export of the most powerful computer chips to China. Such restrictions are less useful—and thus not observed in practice—for chips that China can also manufacture itself. Selective cooperation thus makes sense particularly with regard to cutting-edge technologies; less so for everything else.

Avoiding Global Arms Races/Races to the Bottom

AI also generates prisoners-dilemma-style collective action problems.[34] A particularly vexing variant is the "race to the bottom". Governments might like to restrict the development or deployment of AI to avoid negative side effects also inside their borders. But they might not dare to, fearing that AI restrictions would only give foreign digital competitors a leg up. Current Brussels discourse is full of such worries about "overregulation harming competitiveness".[35]

In a regulatory race to the bottom, countries may neglect AI safety and human rights protections in the hope that such forbearance lets local companies thrive.[36] The consequences of such competitive laxity can be dire: privacy infringements; premature release of biased systems; an unrestrained push for artificial general intelligence; and insufficient oversight of AI companies and the systems they are building. Countries must cooperate lest they outbid each other in undesirable regulatory laxity.[37]

The most salient collective action problem concerns AI-powered warfare.[38] With respect to LAWS, the US National Security Commission on AI (NSCAI) has explicitly argued that the United States cannot afford unilaterally to forgo such systems, given that enemy countries do not do

so, either—a classic security dilemma.[39] That includes diffusion of AI systems to malignant or authoritarian actors: governments might favor a slower pace of system development or more restricted availability. The felt need to push ahead against competitors tempers such instincts, however, as does the sense that overly cautious rules only benefit less scrupulous AI providers elsewhere.

For deleterious races to the bottom, the key is that cooperation is co-extensive with the space in which products are traded—if AI systems from China were completely banned in Europe, for example, they would also not exert competitive pressure and incite a race to the bottom. In contrast, to obstruct the proliferation of harmful AI, only global agreements are useful. If even one country starts to develop AI-powered weapons, other countries have an incentive to follow suit.

Preventing the Proliferation of Harmful AI

In theory, AI can be used to inflict various kinds of harm, for example through cyber-sabotage, easier access to destructive weapons, tools for espionage, and so on. Observers have therefore worried that harmful AI might end up in the wrong hands—for example terrorists, geopolitical enemies or malignant political rulers, bent on oppressing local populations.[40]

Depending on the non-state actors in question, countries developing AI might be more or less worried that they themselves will become targets of malevolent use. Preventing the proliferation of harmful AI therefore requires pro-active collaboration, in which AI producers commit to restricting technology diffusion in ways that go beyond their narrow self-interest. As is true in other non-proliferation regimes, we can only expect stable and effective international arrangements if all major AI powers are involved. Global cooperation is a *sine qua non* to stem harmful AI diffusion effectively.

Realizing Benefits from Trade

The enormous economies of scale in digital products and hence the low product unit costs mean that *ceteris paribus*, global trade should be welfare-enhancing. At the same time, large companies frequently benefit from the resulting oligopolistic tendencies and amplify them by consciously limiting consumer choice, for example by locking consumers into proprietary digital ecosystems.[41] Big Tech exploitation of dominant market positions has triggered a wide backlash—and thus also dented unrestricted trade in AI products.[42]

In consequence, the relationship between Big Tech dominance and rule-based trade in AI products is more complicated than traditional free trade debates would suggest. At present, a small number of jurisdictions—above all the US and China, but also the EU—have embraced a competitive race to develop and promote home-grown AI.[43]

International agreements could help attenuate these competitive dynamics, as they did, for example, in finance.[44] Managed AI trade would create a sort of truce in global AI development and liberate governments from the current tight embrace with domestic AI champions, where it exists. “Realizing the benefits from trade” would then not only be about profiting from the enormous economies of scale digital products offer, but also about offering governments breathing space and ensuring that technological benefits can be shared broadly.[45]

Inversely, domestic oligopolies might still recreate the undesirable effects of current cross-border ones (what Haggart and Tusikov call Digital Economic Nationalism).[46] Moreover, governments, citizens, and consumers across the world already are customers of the largest AI-supplying companies, but then for products that originally did not rely on AI—think of Google (with Maps, search, Google Docs, etc.), Microsoft (Windows, Office, Azure), Amazon (retail), Apple (devices), and so on. As these companies intertwine AI with their other products, they make it difficult to sever AI trade ties selectively. In short, “limiting trade in AI” is more

complex, and more difficult, than a simple assessment of who benefits and who does not would suggest.

Trade agreements could then help to ensure that foreign AI products and services conform to domestic standards and norms.[47] When trade agreements counter oligopolistic tendencies and ensure that trade benefits are shared broadly, both selective and global cooperation can clearly be beneficial. Selective cooperation on trade can often be more easily achieved, but fairly managed global trade in principle has most to offer.

Sharing Knowledge and Best Practices

Knowledge sharing in the AI field has two contradictory faces. The more governments see AI through a competitive lens, the less willing they may be to share expertise: why help others catch up, or get further ahead? At the same time, publicly sponsored academic research is widely accessible, also to competitors. And many forms of knowledge have limited competitive implications: innovations to make model training and inference less energy consuming or theoretical knowledge about new model architectures that are of little use without access to the right kind of data or computational resources, for example.[48] The competitiveness implications of AI-relevant knowledge, in other words, vary widely, and so do the incentives to cooperate to share it on a reciprocal basis.

It is difficult for governments to develop effective AI regulation, not least when it comes to the technical details, like safety assessments and bias identification—issues with which AI companies themselves still struggle.[49] And even when the latter do claim that their systems are unbiased and transparent, at present governments have limited means for checking compliance with such principles, and therefore for enforcing rules around the unbiasedness, transparency, and explainability of AI systems. That opens scope for cooperation, especially

for countries with limited homegrown AI regulatory capabilities or capacities.[50] Thailand, for example, cooperates with Germany and the United States, among others, on the governance and adoption of AI.[51] In instances in which best practice sharing is useful, it ideally happens globally. Where that is politically infeasible, selective cooperation in this domain is still useful.

Promoting Norms

Democratic countries frequently warn against the dangers that AI poses to democracy, as they see it, and they push for international agreements to limit AI-powered mass surveillance, political repression, or the diffusion of political propaganda. The OECD, for example, with mostly functioning democracies as members, has been busy proselytizing for its AI Principles in the Arab Region, Africa, and South America.[52] Human rights protection and the promotion of democratic values in AI use and development are recurrent themes. The EU, too, hopes to diffuse AI Act-style rules to third countries.[53]

The motivations behind these initiatives can differ: democratic governments might well want to support people beyond their borders. They might also want to contain the further spread of digital authoritarianism and the potential loss of geopolitical allies or simply externalize their own rule sets for economic advantage.[54] Here, we remain agnostic about what fuels the norm diffusion initiatives—how well-meaning or strategic they are. But international cooperation could, in theory, foster international alignment regarding AI ethics and promote consensus in this field. In norm promotion, both global and selective cooperation are very useful. If norms regarding ethical AI, for example, put at least some limits on the development and deployment of AI that violates human rights, every single person in any country who is covered by such norms counts as a "win." Thus, even smaller-scale, selective cooperation is valuable for promoting norms (and perhaps more realistic than global cooperation on ethical AI standards, which is very difficult to achieve.[55]

Managing Externalities of Harmful AI and Collective Action Problems

Irrespective of the global dimension, much of the AI governance debate has concentrated on the specific kinds of harms AI systems may do.[56] These harms from AI are thus a specific challenge this cluster of technologies constitutes, beyond the more generic ones such as sharing knowledge or avoiding regulatory races to the bottom.[57]

Importantly for us, these AI harms can travel across borders. For example, AI's environmental and climate impact can be enormous. Companies have frequently trained algorithms on copyrighted materials, also from abroad.[58] Democracy can suffer under the impact of AI-powered fake news and propaganda, and AI disrupts labor markets.[59] Both dynamics can cut across borders, creating incentives for international cooperation, as those affected by the negative physical or policy externalities of other countries will want to cooperate with those countries to minimize externalities.[60]

When it comes to addressing physical and policy externalities, global cooperation is most effective but selective cooperation can be useful as well. When the scope of the externality is itself not global, regional cooperation can be effective. However, even for global externalities, the scope of cooperation need not necessarily be global—agreements even among a smaller number of countries can attenuate competitive dynamics and allow them to pursue sustainability, for example.[61] As is true for avoiding damaging races to the bottom (discussed above), the key is that cooperation is co-extensive with the space in which products are traded. That said, governing externalities stemming from AI harms is a special category, because avoiding them may require interventions in either the development of AI technologies, or in their deployment. Development and deployment in turn have different implications for which kinds of cooperation are useful—an issue that we analyze in detail in the next section.

For now, Table 1 summarizes which kinds of cooperation are particularly useful for which kinds of governance challenges. For example, the proliferation of harmful AI truly requires global cooperation to ensure that there are simply no suppliers for those seeking to procure harmful AI. Benefits from trade, on the other hand, can be realized through selective cooperation among a few countries as well as through global cooperation. In short, the notion that only like-minded countries can (or should) cooperate on AI governance is just as misguided as the notion that AI governance needs to be global. Depending on the goals cooperation is meant to advance, global cooperation can be more useful than selective cooperation (or even the only viable option, e.g., when trying to forestall the proliferation of harmful AI to malicious actors); selective and global cooperation can be equally useful; or selective cooperation can be more useful. Crucially, there is no reason that countries cannot cooperate on some aspects of AI governance and not on others.

	Selective cooperation	*Global cooperation*
Building geopolitical alliances	++	-
Preventing proliferation of harmful AI	-	++
Realizing benefits from trade	++	++
Sharing knowledge and best practices	+	++
Promoting norms	++	++
Managing externalities of harmful AI and collective action problems	+	++

Table 1. Utility of selective versus global cooperation in AI governance. (++ = very useful, + = somewhat useful, - = not useful)

Externalities of Harmful AI Across Lifecycle Phases and Types of AI

Many technologies can generate harms that travel across borders, for example in the form of environmental damage. How those can be tackled effectively depends on the technologies in question and the specific harms they create. We therefore zoom in onto the last row of Table 1—managing externalities of harmful AI and collective action problems—to understand the governance challenges and scope of required cooperation in more detail. To do so, we add two axes of distinction: different phases of the AI lifecycle, and different types of AI.

Development versus Deployment

For our purposes, the AI lifecycle can be usefully disaggregated into two main different phases: system development and system deployment.[62] In principle, these are preceded by a project design phase, which sets downstream constraints by delineating an AI system's task and by collecting, analyzing, and preprocessing the relevant data.[63] However, when it comes to foundation models, for which there are sundry and unforeseen use cases, it is practically impossible to set downstream constraints ex ante.[64]

For us, then, the development phase includes the selection of an appropriate algorithm, tested on the dataset and fine-tuned and validated along the way. During the deployment phase, the trained model is integrated into a target production system. Key goals during this phase include ensuring that the model can be effectively scaled to different workloads and datasets, training the AI system's userbase on how to operate the model, monitoring the model and, if needed, updating it.

These two phases raise different governance issues, so that the salience of the policy concerns identified above varies throughout the AI lifecycle. In the development phase, questions about the data used to train the algorithm and about the unbiasedness, transparency, and fairness of the system being built are particularly pressing.[65] During the deployment phase, concerns about use cases weigh more heavily: is the AI system in question deployed in high-risk scenarios (for example, to make decisions about financial loans or job offers), to enable and scale political repression, or even for criminal ends such as financial malfeasance or terrorism?[66] Managing such use cases is an important focus for policymakers.[67] In contrast to the development phase, however, jurisdictions are in principle able to legislate what can or cannot legally be done with AI systems within their borders—in other words, they can steer their deployment.

Governments have different degrees of leverage over the development and deployment phases of the AI life cycle. As cutting-edge AI originates from a small number of places, most countries will have little leverage over AI development. In principle, governments can legislate what kinds of data companies may use to train their AI systems—forbidding the domestic use of, say, generative AI that was trained on copyrighted material. Given the limited number of model providers, however, the power host governments have over AI development remains limited. Foreign companies may decide to ignore local requirements and simply withhold products from markets altogether if the regulatory strictures do not suit them. Even where they do agree to host country rules, compliance may be hard to monitor. Aligning requirements with countries that actually develop AI then makes sense—if and when it is achievable.

In contrast, governments should have more leverage over AI deployment. When the users of AI systems are large organizations, like big companies or public bodies, governments can regulate use cases relatively easily and thereby steer how AI is deployed within their

jurisdiction. That becomes more difficult when individuals or small organizations are the end users, because in some instances they can access AI-powered services provided from abroad unmonitored, for example software to make illegal deep-fake videos or facial recognition software, easily accessible with a VPN client. Table 2 gives examples of AI applications that fall into the different categories.

	Harm *with* cross-border effects, incentivizing cooperation	Harms *without* cross-border effects, incentivizing *no* cooperation
AI *development*	Development of foundation models that might pose existential risk	Development of biased algorithms in public administration
AI *deployment*	Climate impact of excessive energy use of AI systems	Use of AI for excessive surveillance of public space

Table 2: Examples of AI development and deployment with varying incentives for cooperation

Governance Challenges Across Types of AI

Having laid out potential motivations for AI governance cooperation across the development and deployment phases, we now examine the different governance issues raised by different kinds of AI and analyze whether they call for selective or global cooperation. After all, AI technologies can be applied across fields and industries, with countless use cases, which makes "cooperation in AI governance" a dubiously broad field.[68] An AI-enabled vacuum cleaner raises very different governance problems than an AI-powered drone for the battlefield.

Some of these problems can best be addressed through selective agreement; others require global cooperation. In the household appliance example, countries would need shared

safeguards against excessive personal data collection as a precondition for cross-border sales. Here, agreements among countries that want to trade with each other suffice, and global cooperation remains optional. In the case of AI-powered drones, in contrast, the goals would be to counter excessive proliferation to third countries and malevolent actors, and to enable drone-developing countries to avoid an arms race. Now, only global agreements really work: if even one country liberally sold AI-powered drones, malicious actors could get their hands on them. Other potential producers might quickly follow suit.

These examples show how different kinds of AI, such as AI-powered physical business-to-business or business-to-consumer products and AI-powered military equipment, incentivize different scopes of interstate cooperation. The systems under the AI heading can be sliced and diced in different ways.[69] Rather than canvassing AI technologies as a whole (a questionable enterprise to begin with), we therefore examine four different kinds of AI as examples that raise divergent concerns and thereby divergent cooperation incentives: foundation models, AI-powered physical products, small-scale AI as a service, and militarily relevant AI. Even these four kinds, however, are not mutually exclusive; foundation models, for example, are also software than can be accessed remotely ("as a service"). And different AI systems and model architectures are increasingly combined in real-world applications (Masley *et al.* 2024). That said, these four categories do cover central AI applications and use cases, also as they feature in contemporary political debate. That gives them analytical leverage as we show just how diverse their implications for international cooperation are.

- Foundation Models

AI foundation models are a class of AI models characterized by their vast size, extensive training data, and general-purpose nature.[70] That distinguishes them from other ML models,

which typically perform narrower tasks such as process optimization, image classification, or trend forecasting.[71]

Foundation models are the basis for specialized downstream applications such as chatbots. This versatility presents large challenges for policymakers.[72] For example, malevolent users might tweak a model to bypass built-in safeguards, say to extract instructions to make explosives. At the same time, as models integrate text, sound, video and other content—so-called multi-modal models—they may inch closer towards artificial general intelligence, capable of perceiving, reasoning, and eventually acting across a broad range of modalities. They are likely still far off. But the potential quandaries are clear and serious enough to warrant dedicated attention—clearly reflected in the series of AI safety summits held in the UK in 2023 and in South Korea in 2024.[73]

What do these characteristics imply for cooperation in governance across the two lifecycle phases? The deployment of foundation models can in principle be regulated unilaterally: when is it allowed to use these kinds of models and when not? (Enforcement of deployment rules may well be difficult, but that swings free from international cooperation.) The development of large models, in contrast, is harder to steer if they are built abroad. Given the enormous cost of training the largest models, AI companies may refuse to customize them to each country's wishes.[74] Meta, for example, has already postponed the European roll-out of some AI features because of doubts about its models' compliance with EU rules. If countries were to cooperate and confront foundation model builders with a single set of demands, the latter would be much more likely simply to heed them in model development.

For foundation models, which can be accessed remotely, substantial divergence between domestic AI development rules and those abroad matters a lot. Most governments may face the simple choice of buying what is produced abroad or forgoing them altogether (Mügge

2024b). Differences in the way such models are used at home and abroad, on the other hand, matter less, as unpalatable foreign use cases (e.g., the use of AI-powered domestic surveillance) do not directly create externalities for other countries.

- AI-Powered Physical Products

Consider next a completely different category of AI systems: physical products whose functionality depends on AI—modern cars or kitchen appliances, for example, but also medical or communication devices that may malfunction when you really cannot afford it. Here safety clearly is an issue, as is resilience against cyberattacks and potentially excessive data collection by device manufacturers.

How do cooperation incentives look for this category of AI systems? As was true for foundation models, jurisdictions can in principle set local rules for system deployment—the conditions under which AI-powered hardware may be used. In contrast to the foundation models, however, these devices typically come with simpler systems built into them. The software powering them may still be complex, but nowhere near as complex as foundation models. That makes it possible for companies to finetune systems to local demands at relatively little cost—think of integrated privacy safeguards, for example. In consequence, for these types of products, the incentive to cooperate is lower in the development phase than for foundation models.

For AI-powered physical products, it matters relatively little whether domestic development and deployment rules diverge substantially from those abroad. After all, governments can insist that foreign products meet local specifications (which then need to be “programmed” into the products in the development phase) and can stop them from being imported if they fail to do so. Even then, this use of restrictions is especially relevant for high-risk products (e.g., autonomous vehicles and AI-powered medical devices). In the case of low-risk products (e.g.,

wearable fitness devices and smart home appliances), domestic authorities will more often be willing simply to follow reasonable home country rules, given the potential extra costs of imposing an extra set of rules on foreign producers. For those products, cooperation might de facto take the form of adoption of home country standards.

- Small-scale AI as a service

AI is integrated into many different kinds of services that are available online. In a crucial difference to the previous category, the products in question are not built into specific devices—such as a vacuum cleaner—but can be transferred digitally to the end user or accessed remotely.

This digital character means that it is inherently difficult to keep unwanted software outside a jurisdiction; regulatory interdependence in inevitably higher (Mügge 2024b). Just think of digital personal assistants, recommender systems used for content feed personalization, image recognition software, or websites that offer deepfake videos—all of which may compromise user safety or privacy. Effective control over such applications' availability within a country's borders can require substantial cooperation.

This category partially overlaps with the foundation models discussed above. For our purposes, however, we draw a distinction between large models, owned and operated by a small number of highly visible companies, and relatively more simple systems that may be offered by a wide range of companies, like tools for voice cloning or other deepfakes. In this category, the incentives for cooperation are different from those discussed so far. The problem is not mandating what AI tools could or could not look like (the development phase), but effectively enforcement of how they are made available and used (deployment)—because even if outlawed domestically, citizens could still access them remotely. Here, cooperation thus focuses on restricting the availability of systems in the first place, such that their

deployment can be internationally coordinated. And as was true for the previous category, the cooperation imperative is much higher for high-risk systems (e.g., AI-powered surveillance systems and autonomous trading systems) than for low-risk ones (e.g., customer service chatbots and language translation systems), which domestic authorities may treat with benign neglect.

For small-scale AI as a service, differences between domestic development rules and those abroad do not matter a lot. In principle, the specifications of small-scale AI as a service can easily be changed to conform with the preferences of countries importing these services. However, divergent deployment rules abroad matter because they affect to whom, and under which circumstances, these systems are available.

- Militarily Relevant AI

AI is transforming modern military operations in numerous ways, ranging from autonomous systems to decision-making aids.[75] Military AI applications fall into three categories: sustainment and support refers to AI systems that enhance logistics, cybersecurity, and operational efficiencies.[76] For example, AI aids in managing vast data flows, securing communication networks, and improving resource distribution. Adversarial and non-kinetic uses include AI-driven systems designed to counter cyber-attacks, conduct surveillance, and perform offensive cyber operations without physical combat. In contrast, adversarial and kinetic uses involve more direct engagement, such as AI-assisted targeting systems, autonomous weapon platforms, and combat simulations, potentially reducing human involvement in battlefield decision-making. While AI enhances operational capabilities, its integration into military systems introduces ethical quandaries, such as loss of human oversight, escalation of conflict, and issues of accountability and transparency.[77]

These concerns incentivize governments to cooperate in order to try to establish international frameworks that regulate both the development and deployment of AI-powered military equipment in international contexts, akin to international rules of war more generally. One key challenge for such cooperation is the fact that once one country starts developing AI-powered weapons, a security dilemma ensues in which other countries are tempted to develop similar weapons in response.[78] Hence the only form of cooperation that works here is truly global cooperation, including at least all major military powers. However, such global cooperation is difficult to achieve, as each country individually has an incentive not to cooperate even if all countries would be better off if everyone did—a classic prisoners dilemma.

		Divergence between domestic AI *development* rules and those abroad…	
		…matters a lot, making cooperation on AI development desirable.	…does not matter a lot, making cooperation on AI development optional.
Divergence between domestic AI *deployment* rules and those abroad…	…matters a lot, making cooperation on AI development desirable.	Military AI	Small-scale AI as a service
	…does not matter a lot, making cooperation on AI development optional.	Foundation models	AI-powered physical products

Table 3. Cross-border impact of domestic regimes for AI governance

When it comes to militarily relevant AI, it matters a lot whether domestic development and deployment rules differ from those abroad. Foreign development rules determine whether AI-powered weapons, for example, are fully autonomous or keep a human in the loop in a meaningful way. This makes a difference in warfare.[79] And since militarily relevant AI is not

usually meant to be deployed locally (unless there is a civil war or government repression), the deployment rules of countries that produce militarily relevant AI have serious, tangible consequences for other countries. Table 3 summarizes the cross-border impact of domestic regimes governing the development and deployment of the four different kinds of AI identified above.

Having identified four different kinds of AI that engender different policy concerns, we can now map them onto the different phases of the AI lifecycle to see which concerns are most pressing in which phase and build a systematic understanding of when and how countries might want to cooperate to address these concerns effectively. Table 4 below integrates our previous arguments about how these four exemplary kinds of AI do or do not create incentives for cooperation across the AI lifecycle.

	Unilateral regulation		*Selective cooperation*		*Global cooperation*	
	Development	*Deployment*	*Development*	*Deployment*	*Development*	*Deployment*
Foundation models	-	+	+	-	++	-
AI-powered physical products	-	+	+	-	++	-
Small-scale software as a service	-	+	+	-	++	-
Militarily relevant AI	-	-	+	+	++	++

Table 4. Relative benefits of different scopes of cooperation to avoid harms specific to AI types, split across development and deployment phases of the AI life cycle (++ = very useful, + = somewhat useful, - = not useful)

Conclusion

Just about any expert in the field agrees that there is something to be won in international cooperation on AI. But how, and with whom? To our mind, too much of the debate falls into clear camps: either cooperation should be global (which normally means including Europe, the United States, and certainly China, too), or it should be selective, involving “like-minded countries” (so excluding China).

Our analysis suggests that that view is too simplistic. Incentives for cooperation vary, and they may point in different directions. For some aspects of AI policy, global cooperation is the only approach that works; for others, selective cooperation is a useful second-best or even an equally worthwhile solution. That means that governments should not approach global and selective cooperation as either/or alternatives. Instead, a variable geometry of cooperation on AI matters is not only possible, but indeed a desirable approach to such a broad field with such diverse concerns.

On the one hand, then, the “we need global cooperation”-camp should realize that, desirable as such encompassing agreements would be, more circumscribed cooperations may deserve their support, as well. If those are the only constellations that work, then so be it.

On the other hand, and more importantly to our mind, there is no reason why AI cooperation should completely follow an alliance-building logic. The contemporary geo-economic competition between China and the United States is a political reality. It would take more than successful AI cooperation to supersede that antagonism. But that does not mean that there would not be many areas in which cooperation could be beneficial to both sides—think of avoiding arms races, forestalling the unchecked proliferation of dangerous AI to terrorists and

other criminals, creating basic safety standards for AI-powered consumer appliances, and so on. Several cooperation logics can exist side by side.

By the same token, there is no reason why the EU—while clearly a longstanding ally of the United States—should slavishly follow the American lead on AI cooperation. There are forms of knowledge sharing, trade in non-sensitive products, ethics discussions, safety research and so on that could be beneficially pursued along a Sino-European axis even if the United States were not involved.

For further research, it would be fruitful to analyze whether and how these cooperation logics play out in practice. There is lively debate about the degree to which the EU and the United States should align AI governance with each other and how much scope there is for cooperation with China. But how do these three powers (and others) actually behave? There is a multitude of different AI governance initiatives underway in global and regional forums. Examining which of the imperatives we have outlined here actually prevail in practice, and why, will be a worthwhile next step.

Endnotes

[1] UN High-Level Advisory Body on Artificial Intelligence 2023.
[2] Büthe and Mattli 2011; Hoos and Irgens 2023.
[3] UN High-Level Advisory Body on Artificial Intelligence 2023, p.14-16. It is noteworthy, however, that in its final report, the UN AI Advisory Body walked back from this recommendation and instead suggested a "light-touch" AI Office within the UN Secretariat that would serve as a coordinating mechanism for existing AI governance initiatives (United Nations High-Level Advisory Body on Artificial Intelligence 2024).
[4] Global Partners Digital 2024; Kerry 2024; Kerry et al. 2024.
[5] Cordell 2020; Allen 2020.
[6] von Ingersleben-Seip 2023.
[7] Mueller 2024; Nitzberg and Zysman 2022; Taeihagh 2021; cf. Beaumier et al. 2020.
[8] Justo-Hanani 2022.
[9] Ostrom 2009.
[10] Mitchell 2019; Hitzler et al. 2022. The term "symbolic" here refers to a specific approach in AI that involves representing knowledge and reasoning using symbols, rules, and logic. This approach contrasts with machine learning, which focuses more on data-driven learning and statistical patterns.
[11] International Telecommunication Union (ITU) 2024.
[12] Searle 1995.
[13] von Ingersleben-Seip 2023; Ulnicane 2024.
[14] Organization for Economic Cooperation and Development (OECD) 2024.
[15] North Atlantic Treaty Organization (NATO) 2024; Ministry of Foreign Affairs of Japan 2023; Council of Europe 2024; cf. Hoch and Mügge 2025; Indian Ministry of Electronics & IT 2024.
[16] U.S. Department of State n.d. Even though, as of September 2025, it is unclear whether or in which form the EU-US Trade and Technology Council (TTC) will continue to exist.
[17] Echikson *et al.* 2024.
[18] UN High-Level Advisory Body on Artificial Intelligence 2023.
[19] Yi 2024.
[20] Ala-Pietilä and Smuha 2021, p. 237.
[21] Kerry *et al.* 2024; Cheng & Zeng 2023.
[22] Cf. Paul 2023.
[23] White House 2021.
[24] Global Partnership on Artificial Intelligence (GPAI) n.d.
[25] Solomon 2023; Politi 2021.
[26] Dai *et al.* 2010.
[27] Lemke *et al.* 2023.
[28] Waltz 2010.
[29] Keohane 2005.

[30] Wendt 1992.
[31] Drezner 2010.
[32] Katagiri 2024; Borchert et al. 2024.
[33] cf. National Security Commission on Artificial Intelligence 2021.
[34] In game theory, a prisoner's dilemma game is one in which non-cooperation is the indivually rational choice for each of the players even though cooperation would yield higher payoffs for each (e.g., Snidal 1985).
[35] Draghi 2024.
[36] Scholvin & Wigell 2018.
[37] Russell 2019; Tegmark 2017.
[38] Bode & Huelss 2022.
[39] NSCAI 2021.
[40] Kreps 2021.
[41] Open Markets Institute 2023; Nitzberg & Zysman 2022; Srnicek 2017; Staab 2019.
[42] On the backlash, see e.g. Durand 2020.
[43] Bradford 2023; Lee 2018; Mügge 2024a; critically Bryson and Malikova 2021.
[44] Singer 2007.
[45] McKinsey Global Institute n.d.; Korinek & Stiglitz 2019, 2021.
[46] Haggart & Tusikov 2023.
[47] Ulnicane 2022.
[48] Widder et al. 2023.
[49] Balasubramaniam *et al.* 2023; Zuiderveen Borgesius 2018.
[50] Cafaggi & Pistor 2015; Bach & Newman 2007; see also Lavenex *et al.* 2021.
[51] National Electronics and Computer Technology Center 2022.
[52] Russo & Oder 2023.
[53] Almada & Radu 2024.
[54] Bradford 2023.
[55] von Ingersleben-Seip 2023.
[56] Crawford 2021; Dauvergne 2021.
[57] Smuha 2021.
[58] Cf. Haggart & Tusikov 2023.
[59] Bridle 2018; Acemoglu & Johnson 2023; Frey 2019.
[60] Abbott & Snidal 2001.
[61] Ostrom 2009.
[62] Burr & Leslie 2023; De Silva & Alahakoon 2022.
[63] Spector et al. 2022.
[64] Cf. Gebru & Torres 2024.
[65] Gasser & Mayer-Schönberger 2024.
[66] Smith & Browne 2019.
[67] Laux *et al.* 2024, Mittelstadt 2024; Paul 2023.
[68] Brynjolfsson et al. 2021.
[69] Spector *et al*. 2022.
[70] E.g. Moës & Ryan 2023.
[71] Mitchell 2019.

[72] Seger *et al.* 2023.
[73] Cf. Bostrom 2014; Russell 2019; Schneider 2019.
[74] Vipra & Myers West 2023.
[75] Firlej & Taeihagh 2021; Borchert et al. 2024.
[76] Taddeo et al. 2021.
[77] Taddeo et al. 2021.
[78] Booth & Wheeler 2023.
[79] Firlej & Taeihagh 2021.

References

Abbott, Kenneth W., and Duncan Snidal. 2001. 'International "standards" and International Governance'. *Journal of European Public Policy* 8(3): 345–70. doi:10.1080/13501760110056013.

Acemoglu, Daron, and Simon Johnson. 2023. *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. New York: PublicAffairs.

Ala-Pietilä, Pekka, and Nathalie A. Smuha. 2021. 'A Framework for Global Cooperation on Artificial Intelligence and Its Governance'. In *Reflections on Artificial Intelligence for Humanity*, eds. Bertrand Braunschweig and Malik Ghallab. Cham: Springer International Publishing, 237–65. doi:10.1007/978-3-030-69128-8_15.

Allen, Gregory C. 2022. 'One Key Challenge for Diplomacy on AI: China's Military Does Not Want to Talk'. https://www.csis.org/analysis/one-key-challenge-diplomacy-ai-chinas-military-does-not-want-talk (March 27, 2024).

Almada, Marco, and Anca Radu. 2024. 'The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy'. *German Law Journal*: 1–18. doi:10.1017/glj.2023.108.

Bach, David, and Abraham L. Newman. 2007. 'The European Regulatory State and Global Public Policy: Micro-Institutions, Macro-Influence'. *Journal of European Public Policy*. doi:10.1080/13501760701497659.

Balasubramaniam, Nagadivya, Marjo Kauppinen, Antti Rannisto, Kari Hiekkanen, and Sari Kujala. 2023. 'Transparency and Explainability of AI Systems: From Ethical Guidelines to Requirements'. *Information and Software Technology* 159: 107197. doi:10.1016/j.infsof.2023.107197.

Beaumier, Guillaume, Kevin Kalomeni, Malcolm Campbell-Verduyn, Marc Lenglet, Serena Natile, Marielle Papin, Daivi Rodima-Taylor, Arthur Silve, and Falin Zhang. 2020. 'Global Regulations for a Digital Economy: Between New and Old Challenges'. *Global Policy* 11(4): 515–22. doi:10.1111/1758-5899.12823.

Bode, Ingvild, and Hendrik Huelss. 2022. *Autonomous Weapons Systems and International Norms*. Montreal: McGill University Press.

Booth, Ken, and Nicholas J. Wheeler. 2023. 'Uncertainty'. In *Security Studies*, Routledge.

Borchert, Heiko, Torben Schütz, and Jospeh Verbovsky. 2024. *The Very Long Game. 25 Case Studies on the Global State of Defense AI*. Cham: Springer Nature.

Bostrom, Nick. 2014. *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bradford, Anu. 2023. *Digital Empires: The Global Battle to Regulate Technology*. Oxford University Press.

Bridle, James. 2018. *New Dark Age. Technology and the End of the Future*. London: Verso.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2021. 'The Productivity J-Curve: How Intangibles Complement General Purpose Technologies'. *American Economic Journal: Macroeconomics* 13(1): 333–72. doi:10.1257/mac.20180386.

Burr, Christopher, and David Leslie. 2023. 'Ethical Assurance: A Practical Approach to the Responsible Design, Development, and Deployment of Data-Driven Technologies'. *AI and Ethics* 3(1): 73–98. doi:10.1007/s43681-022-00178-0.

Büthe, Tim, and Walter Mattli. 2011. 'The New Global Rulers: The Privatization of Regulation in the World Economy'. In *The New Global Rulers*, Princeton University Press. doi:10.1515/9781400838790.

Cafaggi, Fabrizio, and Katharina Pistor. 2015. 'Regulatory Capabilities: A Normative Framework for Assessing the Distributional Effects of Regulation'. *Regulation & Governance* 9(2): 95–107. doi:10.1111/rego.12065.

Cheng, Jing, and Jinghan Zeng. 2023. 'Shaping AI's Future? China in Global AI Governance'. *Journal of Contemporary China* 32(143): 794–810. doi:10.1080/10670564.2022.2107391.

Cordell, Kristen. 2020. 'The International Telecommunication Union: The Most Important UN Agency You Have Never Heard Of'. https://www.csis.org/analysis/international-telecommunication-union-most-important-un-agency-you-have-never-heard (March 27, 2024).

Council of Europe. 2024. 'The Framework Convention on Artificial Intelligence - Artificial Intelligence - Www.Coe.Int'. *Artificial Intelligence*. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence (November 11, 2024).

Crawford, Kate. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Dai, Xinyuan, Duncan Snidal, and Michael Sampson. 2010. 'International Cooperation Theory and International Institutions'. In *Oxford Research Encyclopedia of International Studies*, doi:10.1093/acrefore/9780190846626.013.93.

Dauvergne, Peter. 2021. 'The Globalization of Artificial Intelligence: Consequences for the Politics of Environmentalism'. *Globalizations* 18(2): 285–99. doi:10.1080/14747731.2020.1785670.

De Silva, Daswin, and Damminda Alahakoon. 2022. 'An Artificial Intelligence Life Cycle: From Conception to Production'. *Patterns* 3(6): 100489. doi:10.1016/j.patter.2022.100489.

Draghi, Mario. 2024. *The Future of European Competitiveness*. Luxembourg: Publications Office of the European Union.

Drezner, Daniel W. 2010. 'Mercantilist and Realist Perspectives on the Global Political Economy'. In *Oxford Research Encyclopedia of International Studies*, doi:10.1093/acrefore/9780190846626.013.260.

Durand, Cédric. 2020. *Techno-Féodalisme - Critique de l'économie Numérique*. Paris: Zones. La Découverte.

Echikson, William, Dimitar Lilkov, and Clara Riedenstein. 2024. 'Transatlantic Trade and Technology: Partners or Rivals?' *CEPA*. https://cepa.org/comprehensive-reports/transatlantic-trade-and-technology-partners-or-rivals/ (June 17, 2025).

'ETDA to Launch Thailand's First AI Governance Clinic (AIGC) - NECTEC : National Electronics and Computer Technology Center'. 2022. https://www.nectec.or.th/en/about/news/etda-aigc-promote.html (March 27, 2024).

Firlej, Mikolaj, and Araz Taeihagh. 2021. 'Regulating Human Control over Autonomous Systems'. *Regulation & Governance* 15(4): 1071–91. doi:10.1111/rego.12344.

Frey, Carl Benedict. 2019. *The Technology Trap: Capital, Labor, and Power in the Age of Automation*. Princeton: Princeton University Press.

Gasser, Urs, and Viktor Mayer-Schönberger. 2024. *Guardrails. Guiding Human Decisions in the Age of AI*. Princeton: Princeton University Press.

Gebru, Timnit, and Émile P. Torres. 2024. 'The TESCREAL Bundle: Eugenics and the Promise of Utopia through Artificial General Intelligence'. *First Monday* 29(4). doi:10.5210/fm.v29i4.13636.

Global Partners Digital. 2024. 'The Final Report of HLAB-AI: Our Analysis and Thoughts – Global Partners Digital'. https://www.gp-digital.org/news/the-final-report-of-hlab-ai-our-analysis-and-thoughts/ (April 1, 2025).

Haggart, Blayne, and Natasha Tusikov. 2023. *The New Knowledge. Information, Data and the Remaking of Global Power*. Lanham: Rowman & Littlefield.

Hitzler, Pascal, Aaron Eberhart, Monireh Ebrahimi, Md Kamruzzaman Sarker, and Lu Zhou. 2022. 'Neuro-Symbolic Approaches in Artificial Intelligence'. *National Science Review* 9(6): nwac035. doi:10.1093/nsr/nwac035.

Hoch, Sina and Daniel Mügge, "EU Influence in Global AI Governance and Its Limits", RegulAite working paper 01/2025.

Hoos, Holger, and Morten Irgens. 2023. 'Viewpoint: Europe Needs a CERN for Artificial Intelligence | Science|Business'. https://sciencebusiness.net/viewpoint/ai/viewpoint-europe-needs-cern-artificial-intelligence (April 2, 2025).

House, The White. 2021. 'Remarks by President Biden at the 2021 Virtual Munich Security Conference'. *The White House*. https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/02/19/remarks-by-president-biden-at-the-2021-virtual-munich-security-conference/ (March 27, 2024).

Indian Ministry of Electronics & IT. 2024. '6th Meeting of the GPAI Ministerial Council Held on 3rd July 2024 at New Delhi'. https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=2030534 (August 20, 2024).

von Ingersleben-Seip, Nora. 2023. 'Competition and Cooperation in Artificial Intelligence Standard Setting: Explaining Emergent Patterns'. *Review of Policy Research* 40(5): 781–810. doi:10.1111/ropr.12538.

International Telecommunication Union (ITU). 2024. 'Key Findings on the State of Global AI Governance'. *ITU*. https://www.itu.int/hub/2024/07/key-findings-on-the-state-of-global-ai-governance/ (August 21, 2024).

Justo-Hanani, Ronit. 2022. 'The Politics of Artificial Intelligence Regulation and Governance Reform in the European Union'. *Policy Sciences* 55(1): 137–59. doi:10.1007/s11077-022-09452-8.

Katagiri, Nori. 2024. 'Artificial Intelligence and Cross-Domain Warfare: Balance of Power and Unintended Escalation'. *Global Society* 38(1): 34–48. doi:10.1080/13600826.2023.2248179.

Keohane, Robert O. 2005. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton University Press.

Kerry, Cameron F. 2024. 'The Good, the Not-so-Good, and the Ugly of the UN's Blueprint for AI'. *Brookings*. https://www.brookings.edu/articles/the-good-the-not-so-good-and-the-ugly-of-the-uns-blueprint-for-ai/ (September 17, 2024).

Kerry, Cameron F., Joshua P. Meltzer, Andrea Renda, and Andrew W. Wyckoff. 2024. 'Should the UN Govern Global AI?' *Brookings*. https://www.brookings.edu/articles/should-the-un-govern-global-ai/ (August 19, 2024).

Korinek, Anton, and Joseph E. Stiglitz. 2019. '14. Artificial Intelligence and Its Implications for Income Distribution and Unemployment'. In *14. Artificial Intelligence and Its Implications for Income Distribution and Unemployment*, University of Chicago Press, 349–90. doi:10.7208/9780226613475-016.

Korinek, Anton, and Joseph E. Stiglitz. 2021. 'Artificial Intelligence, Globalization, and Strategies for Economic Development'. doi:10.3386/w28453.

Kreps, Sarah. 2021. *Democratizing Harm: Artificial Intelligence in the Hands of Nonstate Actors*. Washington DC: Brookings Institution.

Laux, Johann, Sandra Wachter, and Brent Mittelstadt. 2024. 'Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk'. *Regulation & Governance* 18(1): 3–32. doi:10.1111/rego.12512.

Lavenex, Sandra, Omar Serrano, and Tim Büthe. 2021. 'Power Transitions and the Rise of the Regulatory State: Global Market Governance in Flux'. *Regulation & Governance* 15(3): 445–71. doi:10.1111/rego.12400.

Lee, Kai-Fu. 2018. *AI Superpowers. China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin.

Lemke, Nicole, Philipp Trein, and Frédéric Varone. 2023. 'Agenda-Setting in Nascent Policy Subsystems: Issue and Instrument Priorities across Venues'. *Policy Sciences* 56(4): 633–55. doi:10.1007/s11077-023-09514-5.

Masley, Nestor, Fattorini, Loredana, Raymond Perrault, Vanessa Parli, Anka Reuel, Erik Brynjolfsson, John Etchemendy, et al. 2024. *The AI Index 2024 Annual Report*. Stanford: Stanford University. https://aiindex.stanford.edu/report/.

'McKinsey Global Institute (MGI)'. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx (March 27, 2024).

Ministry of Foreign Affairs of Japan. 2023. 'G7 Leaders' Statement on the Hiroshima AI Process'. *Ministry of Foreign Affairs of Japan*. https://www.mofa.go.jp/ecm/ec/page5e_000076.html (August 20, 2024).

Mitchell, Melanie. 2019. *Artificial Intelligence. A Guide for Thinking Humans*. London: Pelican Books.

Moës, Nicolas, and Frank Ryan. 2023. *Heavy Is the Head the Wears the Crown*.

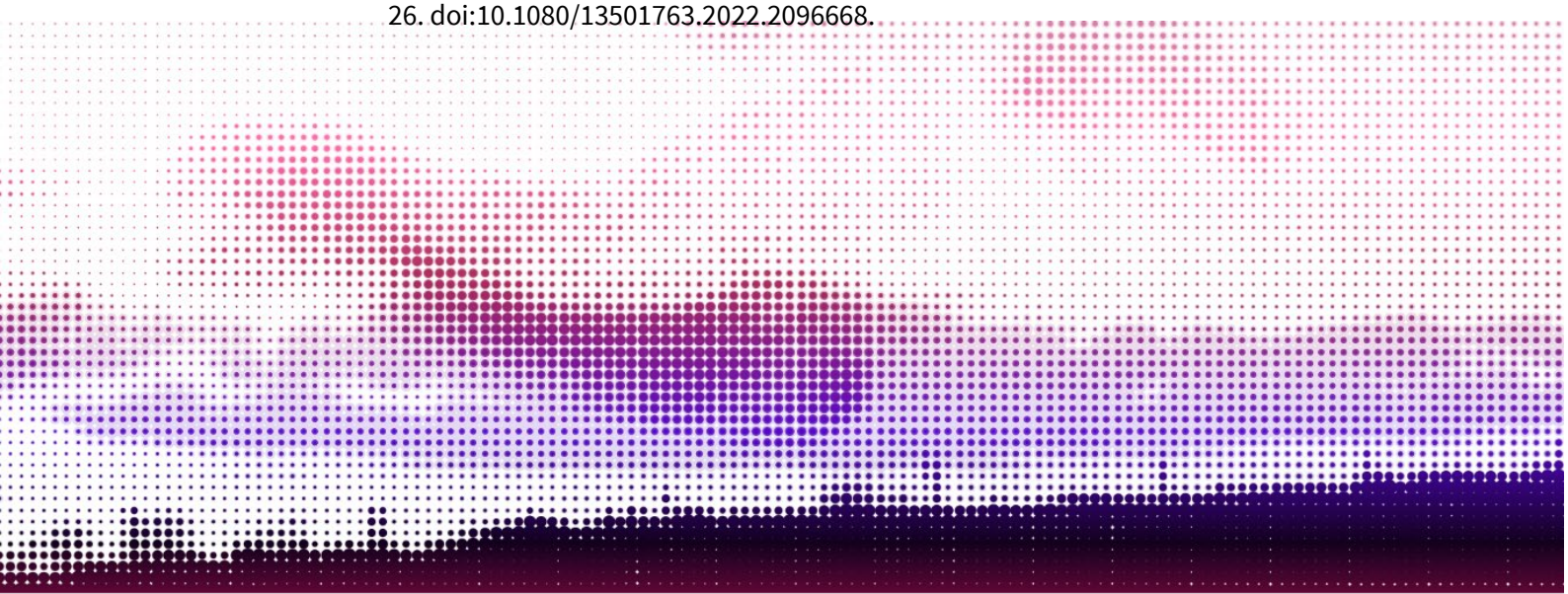
Mueller, Milton. 2024. 'Rethinking AI Governance: The Political Economy of the Digital Ecosystem'. https://papers.ssrn.com/abstract=4911002 (November 11, 2024).

Mügge, Daniel. 2024a. 'EU AI Sovereignty: For Whom, to What End, and to Whose Benefit?' *Journal of European Public Policy* 31(8): 2200–2225. doi:10.1080/13501763.2024.2318475.

Mügge, Daniel. 2024b. 'Regulatory Interdependence in AI'. In *Handbook on Public Policy and Artificial Intelligence*, eds. Regine Paul, Emma Carmel, and Jennifer Cobbe. Cheltenham Spa: Edward Elgar.

National Security Commission on AI. 2021. *Final Report*. Arlington VA.

Nitzberg, Mark, and John Zysman. 2022. 'Algorithms, Data, and Platforms: The Diverse Challenges of Governing AI'. *Journal of European Public Policy*: 1–26. doi:10.1080/13501763.2022.2096668.



North Atlantic Treaty Organization (NATO). 2024. 'Summary of NATO's Revised Artificial Intelligence Strategy'. *NATO*. https://www.nato.int/cps/en/natohq/official_texts_227237.htm (August 20, 2024).

NSCAI. 2021. 'Final Report by the National Security Commission on AI'. https://reports.nscai.gov/final-report/introduction.

Open Markets Institute. 2023. *AI in the Public Interest: Confronting the Monopoly Threat*. Washington DC.

Organization for Economic Cooperation and Development (OECD). 2024. 'OECD Legal Instruments'. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (August 20, 2024).

Ostrom, Elinor. 2009. 'A Polycentric Approach for Coping with Climate Change'. doi:10.2139/ssrn.1934353.

Paul, Regine. 2023. 'European Artificial Intelligence "Trusted throughout the World": Risk-Based Regulation and the Fashioning of a Competitive Common AI Market'. *Regulation & Governance*: rego.12563. doi:10.1111/rego.12563.

Politi, Daniel. 2021. 'U.S. Has Lost AI Race to China, According to Former Software Chief at Pentagon'. *Slate*. https://slate.com/news-and-politics/2021/10/us-lost-artificial-intelligence-race-china-nicolas-chaillan.html (March 27, 2024).

Russell, Stuart. 2019. *Human Compatible. Artificial Intelligence and the Problem of Control*. New York: Viking.

Russo, Lucia, and Noah Oder. 2023. 'How Countries Are Implementing the OECD Principles for Trustworthy AI'. https://oecd.ai/en/wonk/national-policies-2 (August 21, 2024).

Schneider, Susan. 2019. *Artificial You. AI and the Future of Your Mind*. Princeton: Princeton University Press.

Scholvin, Sören, and Mikael Wigell. 2018. 'Power Politics by Economic Means: Geoeconomics as an Analytical Approach and Foreign Policy Practice'. *Comparative Strategy* 37(1): 73–84. doi:10.1080/01495933.2018.1419729.

Searle, John R. 1995. *The Construction of Social Reality*. New York: Free Press.

Singer, David. 2007. *Regulating Capital. Setting Standards for the International Financial System*. Ithaca: Cornell University Press.

Smith, Brad, and Carol Ann Browne. 2019. *Tools and Weapons. The Promise and Peril of the Digital Age*. London: Penguin.

Smuha, Nathalie. 2021. 'Beyond the Individual: Governing AI's Societal Harm'. *Internet Policy Review* 10(3): 1–32.

Snidal, Duncan. 1985. 'Coordination versus Prisoners' Dilemma: Implications for International Cooperation and Regimes'. *American Political Science Review* 79(4): 923–42. doi:10.2307/1956241.

Solomon, Jay. 2023. 'Ex-Pentagon Chief: US, China Locked in Existential Struggle for AI Dominance | Semafor'. https://www.semafor.com/article/07/27/2023/us-china-existential-struggle-ai-dominance (March 27, 2024).

Spector, Alfred, Peter Norvig, Chris Wiggins, and Jeanette Wing. 2022. *Data Science in Context. Foundations, Challenges, Opportunities*. Cambridge: Cambridge University Press.

Srnicek, Nick. 2017. *Platform Capitalism*. Cambridge: Polity Press.

Staab, Philipp. 2019. *Digitaler Kapitalismus. Markt Und Herrschaft in Der Ökonomie Der Unknappheit*. Frankfurt am Main: Suhrkamp.

Taddeo, Mariarosaria, David McNeish, Alexander Blanchard, and Elizabeth Edgar. 2021. 'Ethical Principles for Artificial Intelligence in National Defence'. *Philosophy & Technology* 34(4): 1707–29. doi:10.1007/s13347-021-00482-3.

Taeihagh, Araz. 2021. 'Governance of Artificial Intelligence'. *Policy and Society* 40(2): 137–57. doi:10.1080/14494035.2021.1928377.

Tegmark, Max. 2017. *Life 3.0. Being Human in the Age of AI*. London: Penguin.

Ulnicane, Inga. 2022. 'Artificial Intelligence in the European Union'. In *The Routledge Handbook of European Integrations*, Taylor & Francis. doi:10.4324/9780429262081-19.

Ulnicane, Inga. 2024. 'Governance Fix? Power and Politics in Controversies about Governing Generative AI'. *Policy and Society*: puae022. doi:10.1093/polsoc/puae022.

'UN High-Level Advisory Body on Artificial Intelligence'. https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf (March 27, 2024).

United Nations High-Level Advisory Body on Artificial Intelligence. 2024. *Govering AI for Humanity: Final Report*. New York, NY: United Nations.

U.S. Department of State. 'U.S.-EU Trade and Technology Council (TTC)'. *United States Department of State*. https://www.state.gov/u-s-eu-trade-and-technology-council-ttc/ (August 20, 2024).

Vipra, Jai, and Sarah Myers West. 2023. *Computational Power and AI*. New York. https://ainowinstitute.org/publication/policy/compute-and-ai.

Waltz, Kenneth N. 2010. *Theory of International Politics*. Waveland Press.

Wendt, Alexander. 1992. 'Anarchy Is What States Make of It: The Social Construction of Power Politics'. *International Organization* 46(2): 391–425.

Widder, David Gray, Sarah West, and Meredith Whittaker. 2023. 'Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI'. doi:10.2139/ssrn.4543807.

Yi, Wang. 2024. 'Wang Yi on Global AI Governance: Ensure That AI Is a Force for Good, Ensure Safety and Ensure Fairness'. https://www.mfa.gov.cn/mfa_eng/wjbzhd/202403/t20240308_11256430.html (March 27, 2024).

Zuiderveen Borgesius, Frederik. 2018. *Discrimination, Artificial Intelligence, and Algorithmic Decision-Making*. DG Democracy, Council of Europe.