Fragmentation and the Future: Investigating Architectures for International AI Governance

Peter Cihon

University of Oxford

Matthijs M. Maas

University of Cambridge and University of Copenhagen

Luke Kemp

University of Cambridge

Research Article

Abstract

The international governance of artificial intelligence (AI) is at a crossroads: should it remain fragmented or be centralised? We draw on the history of environment, trade, and security regimes to identify advantages and disadvantages in centralising AI governance. Some considerations, such as efficiency and political power, speak for centralisation. The risk of creating a slow and brittle institution, and the difficulty of pairing deep rules with adequate participation, speak against it. Other considerations depend on the specific design. A centralised body may be able to deter forum shopping and ensure policy coordination. However, forum shopping can be beneficial, and fragmented institutions could self-organise. In sum, these trade-offs should inform development of the AI governance architecture, which is only now emerging. We apply the trade-offs to the case of the potential development of high-level machine intelligence. We conclude with two recommendations. First, the outcome will depend on the exact design of a central institution. A well-designed centralised regime covering a set of coherent issues could be beneficial. But locking-in an inadequate structure may pose a fate worse than fragmentation. Second, fragmentation will likely persist for now. The developing landscape should be monitored to see if it is self-organising or simply inadequate.

Policy Implications

- Secretariats of emerging AI initiatives, for example, the OECD AI Policy Observatory, Global Partnership on AI, the UN High-level Panel on Digital Cooperation, and the UN System Chief Executives Board (CEB) should coordinate to halt and reduce further regime fragmentation.
- There is an important role for academia to play in providing objective monitoring and assessment of the emerging AI regime complex to assess its conflict, coordination, and catalysts to address governance gaps without vested interests. Secretariats of emerging AI initiatives should be similarly empowered to monitor the emerging regime. The CEB appears particularly well placed and mandated to address this challenge, but other options exist.
- What AI issues and applications need to be tackled in tandem is an open question on which the centralisation debate sensitively turns. We encourage scholars across AI issues from privacy to military applications to organise venues to more closely consider this vital question.
- Non-state actors, especially those with technical expertise, will have a potent influence in either a fragmented or centralised regime. These contributions need to be used, but there also need to be safeguards in place against regulatory capture.
- The AI regime complex is at an embryonic stage, where informed interventions may be expected to have an outsized impact. The effect of academics as norm entrepreneurs should not be underestimated at this point.

Global Policy (2020) 11:5 doi: 10.1111/1758-5899.12890

AI has the potential to dramatically alter the world for good or ill. These high stakes have driven a recent flurry of international AI policy making at the OECD, G7, G20, and multiple UN institutions. Scholarship has not kept pace with diplomacy. AI governance research to date has predominantly focused on national and sub-national levels (Calo, 2017). AI global governance research remains relatively nascent, focusing mostly on the proliferation of AI ethics principles (Jobin *et al*., 2019) and stocktaking of ongoing initiatives (Garcia, 2020; Schiff *et al*., 2020). Kemp *et al*. (2019) have called for specialised, centralised intergovernmental agencies to coordinate policy responses globally. Others have called for a centralised 'International Artificial Intelligence Organisation' (Erdelyi and Goldsmith, 2018) or an international coordinating mechanism under the G20 (Jelinek *et al*., 2020). Conversely, some scholars favour more decentralised arrangements based around soft law, global standards, or existing international law instruments or UN multilateral organisations (Cihon, 2019; Garcia, 2020; Kunz and Ó hÉigeartaigh, 2020; Wallach and Marchant, 2018).

This paper takes the initial step of considering the question: Should AI governance be centralised? The form of an international regime[1] will fundamentally impact its operation and effectiveness. This includes the critical question of how an institutional form 'fits' the underlying problem (Ekstrom and Crona, 2017; Young, 2002). Questions of regime centralisation have occupied scholars and international negotiations for decades. The US diplomat George Kennan (1970) proposed the establishment of an 'International Environmental Agency' as an initial step towards an International Environmental Authority. The vexing question of whether to have a centralised body for environmental governance continued 42 years later during the Rio + 20 negotiations. There remains significant debate as to how much form affects performance and what level of centralisation is preferable, but there is little doubt that it is an important consideration for international regimes (Biermann and Kim, 2020).

Centralisation is also a neglected area of examination for AI governance. The debate over form is in its infancy for AI with a few proposals for centralised regimes in academic literature and submissions to international processes (Jelinek *et al*., 2020; Kemp *et al*., 2019). Yet it seems unlikely that AI will be immune to increasing discussions and eventual political pushes for regime centralisation. Future negotiations over the form of AI governance will benefit immensely from early analysis.

'Centralisation', in this case, refers to the degree to which the coordination, oversight and/or regulation of a set of AI policy issues or technologies are housed under a single institution. Centralisation is relevant for policy makers and academics alike. A recent report by the UN Secretary General lamented the lack of coordination and inclusion among AI-related initiatives (United Nations Secretary-General, 2020). Early research and anticipatory initiatives may sensitively influence the path governance takes (Stilgoe *et al*., 2013). Scholars have a unique opportunity to be norm entrepreneurs and shape the emerging institutions through proactive, rather than retrospective, work on AI governance. The importance of this proactive approach has been emphasised for emerging technologies more broadly (Rayfuse, 2017). Moreover, choices made today may have long-lasting impacts as AI development continues (Cave and Ó hÉigeartaigh, 2019).

In this paper, we explore the advantages and disadvantages of centralisation for AI governance. The defining problems of AI governance are threefold. The first is the political economy challenge and the importance of non-state actors' expertise in AI. The second is the need for anticipatory governance and technological foresight. The third is the variety and range of different AI applications, technologies, and policy problems.

Our analysis hinges on a comparison with international regimes in three other domain areas, which display these core challenges, specifically environment, trade, and security. These three governance domains, while certainly distinct in important ways, are also arguably similar to AI governance across these dimensions: environmental governance invokes complex scientific questions that require technical expertise, has a broad scope encompassing transboundary and transsector effects, and includes a need for anticipation of future trends and impacts. Trade regimes span across a breadth of individual industries, and involve questions of standard-setting. Security and arms control regimes confront high-stakes situations and strategic interests, and a recurring need to 'modernise' regimes to track ongoing technological change. All three governance domains face questions of institutional inequalities. Finally, these regimes have been the subject of a rich literature exploring fragmentation and centralisations.

We first outline the international governance challenges of AI, and review early proposed responses. We then draw on the conceptual frameworks of 'regime fragmentation' (Biermann *et al*., 2009) and 'regime complexes' (Gómez-Mera *et al*., 2020; Orsini *et al*., 2013), and their application to the history of other international regimes, to identify considerations in designing a centralised regime complex for AI. We conclude with two practical recommendations.

The state of AI governance

Whether AI is a single policy area is actively debated. Some claim that AI cannot be cohesively regulated as it is a collection of disparate technologies, with different applications and risk profiles (Stone *et al*., 2016). This is an important but not entirely convincing objection. The technical field has no settled definition for 'AI',[2] thus it is unsurprising that delineating a manageable scope for AI governance is difficult (Schuett, 2019). Yet this challenge is not unique to AI: definitional issues abound in areas such as environment and energy, but have not figured prominently in debates over centralisation. Indeed, energy and environment ministries are common at the domestic level.

There are numerous ways in which a centralised body could be designed for AI governance. For example, a centralised approach could carve out a subset of interlinked AI issues. This could involve focusing on the potentially high-

risk *applications* of AI systems, such as AI-enabled cyberwarfare, the use of natural language processing for information warfare, lethal autonomous weapons systems (LAWS), or high-level machine intelligence (HLMI).[3] Another approach could govern underlying *resource inputs* for AI such as large-scale compute hardware, software libraries, training datasets, or human talent. We are agnostic on the specifics of how centralisation could or should be implemented. We instead focus on the costs and benefits of centralisation in the abstract. The exact advantages and disadvantages of centralisation will vary with institutional design.

Numerous AI issues could benefit from international cooperation. These include the high-risk applications mentioned above. It also encompasses more quotidian uses, such as AI-enabled cybercrime; human health applications; safety and regulation of autonomous vehicles and drones; surveillance, privacy and data-use; and labour automation. This is not an exhaustive list of international AI policy issues.

Global regulation across these issues is currently nascent, fragmented and evolving. OECD members and several other states agreed to a series of AI Principles, which were subsequently adopted by the G20 (OECD, 2020a). The Global Partnership on AI (GPAI) was launched by the G7 and several other states (GPAI, 2020). The fragmented membership in these initiatives is shown in Figure 1. A wide range of UN institutions have begun to undertake some activities on AI (ITU, 2019). These developments are complimented by various treaty amendments, such as incorporating autonomous vehicles into the 1968 Vienna Convention on Road Traffic (Kunz and Ó hÉigeartaigh, 2020) or ongoing negotiations under the Convention on Certain Conventional Weapons (CCW) on LAWS. Private fora may also influence international governance (See Green and Auld, 2017), including the Partnership on AI and IEEE's Ethically Aligned Design initiative. The UN Secretary General intends to establish a multistakeholder advisory body on global AI cooperation (United Nations Secretary-General, 2020). UNESCO, the Council of Europe, and the OECD have similarly convened multistakeholder groups tasked with drafting policy instruments (Council of Europe (COE), 2020; UNESCO, 2020; ; OECD, 2020b).

Whether these initiatives bear fruit, however, remains unclear, as many of the involved international organisations have fragmented membership, were not originally created to address AI issues and lack effective enforcement or compliance mechanisms (see Morin *et al*., 2019). For instance, while the US has endorsed the OECD AI Principles and while it eventually acquiesced to the GPAI, it has remained sceptical of hard, global rules (Delcker, 2020). China, another global frontrunner in AI, is not a member of either body.[4]

How we initially structure international governance can be critical to its long-term success. Fragmentation and centralisation exist across a spectrum. Some fragmentation will always prevail, absent a global government. But the degree to which it prevails is crucial. Our definitions, including for fragmentation and key terms are provided in Table 1. These definitions are by nature normatively loaded. For example, some may find 'decentralisation' to be a positive framing, while others may see 'fragmentation' to possess negative connotations. Recognising this, we use these terms in an analytical manner.

Centralisation criteria: a history of governance trade-offs

We explore a series of considerations for AI governance based on a review of existing scholarship on fragmentation (Biermann and Kim, 2020; Biermann *et al*., 2009; Ostrom, 2010; Zelli and Asselt, 2013). Specifically, political power and efficient participation support centralisation. The breadth vs. depth dilemma, as well as slowness and brittleness support decentralisation. Policy coordination and forum shopping considerations can cut both ways. This list is substantive, not exhaustive, and we intend it to open a discussion of design considerations for the nascent AI regime complex. It is far from the final word. Within each consideration below, we offer definitions, relevant regime histories, and discussion of implications for AI.

Political power

Regimes embody power in their authority over rules, norms, and knowledge beyond states' exclusive control. A more centralised regime sees this power concentrated among fewer institutions. A centralised, powerful architecture is likely to be more influential against competing international organisations and with constituent states (Orsini *et al*., 2013). Most environmental multilateral treaties, as well as UNEP, have faced sustained criticism for being unable to enact strong, effective rules or enforce them. In contrast, the umbrella of the WTO, has strongly enforced norms such as the most-favoured-nation principle (equally treating all WTO member states) have become the bedrock of international trade. Even to the extent of changing the actions of the US due to WTO rulings. The power and trackrecord of the WTO is so formidable that it has created a chilling effect: the fear of colliding with WTO norms and rules has led environmental treaties to actively avoid discussing or deploying trade-related measures (Eckersley, 2004). The power of this centralised body has stretched beyond the domain of trade to mould related issues.

This is an area of high salience for AI. The creators and chief users of AI are 'big tech' companies which are some of the largest firms in the world by market capitalisation and have already had an enormous effect in shaping government policy (Nemitz, 2018) in favour of 'surveillance capitalism' (Zuboff, 2019). This daunting political economy challenge is perhaps the defining characteristic of AI. It seems unlikely that powerful vested economic and military interests in AI will be steered by a plethora of small bodies better than a single, well-resourced and empowered institution.

Political power offers further benefits in governing emerging technologies that are inherently uncertain in both substance and impact. Uncertainty in technology and preferences has been associated with some increased centralisation in regimes (Koremenos *et al*., 2001a). There may also be benefits to housing a foresight capacity within the

Figure 1. Membership in selected international AI policy initiatives.

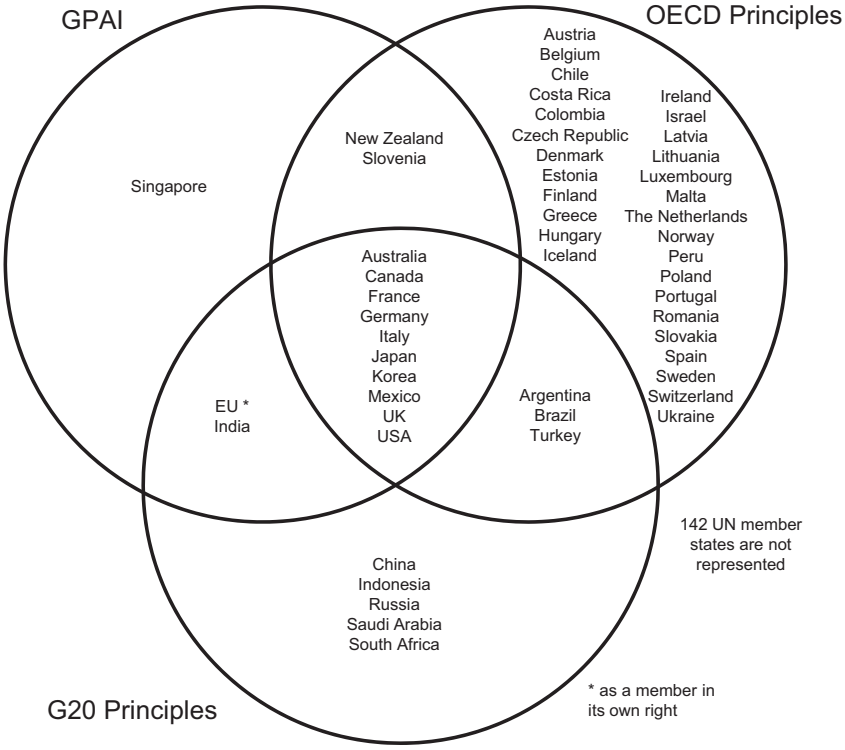

Table 1. Definition of key governance terms

Term	Definition
Fragmentation or decentralisation	A patchwork of international institutions which focus on a particular issue area but differ in scope, membership and often rules (Biermann *et al.*, 2009).
Centralisation	The degree to which governance for an issue lies under the authority of a single body.
Regime complex	A network of three or more international regimes on a common issue area. These should have overlapping membership and cause potentially problematic interactions (Orsini *et al.*, 2013).

regime complex, to allow for accelerated or even proactive efforts (Pauwels, 2019), which would be particularly effective if centralised.

Supporting efficiency and participation

Decentralised AI governance may undermine efficiency and inhibit participation. States often create centralised regimes to reduce costs, for instance by eliminating duplicate efforts, yielding economies of scale within secretariats, and simplifying participation (Esty and Ivanova, 2002). Conversely, fragmented regimes may force states to spread resources and funding over many distinct institutions, limiting the ability of less well-resourced parties to participate (Morin *et al.*, 2019).

Historically, decentralised regimes have presented cost and participation concerns. Hundreds of related and sometimes overlapping international environmental agreements can create 'treaty congestion' (Anton, 2012). This complicates participation and implementation for both developed and developing nations (Esty and Ivanova, 2002). This includes costs associated with travel to different forums, monitoring and reporting for a range of different bodies, and duplication of effort by different secretariats (Esty and Ivanova, 2002). Similar challenges confront decentralised export regimes, which have notable duplication of efforts (Brockmann, 2019).

These challenges are already evident in AI governance. Developing countries are not well represented at most international AI meetings (United Nations Secretary-General, 2020). Simultaneous and globally distributed meetings pose burdensome participation costs. Fragmented organisations must duplicatively invest in high-demand machine learning subject-matter experts to inform their activities. Centralisation would support institutional efficiency and participation.

The costs and participation challenges posed by decentralisation may pose particular barriers to non-state actors (Drezner, 2009). AI-related expertise is primarily located in non-state actors today, namely multinational corporations and universities. Thus, barriers to non-state-actor participation in AI governance will pose particularly acute problems for writing rules that reflect the nature and development trajectory of AI technologies. However, these barriers may not limit all non-state actors from engaging in multiple fora. Indeed, those with sufficient resources may be able to pursue strategies to their advantage (Kuyper, 2014).

Slowness and brittleness of centralised regimes

One problem of centralisation lies in the relatively slow process of establishing centralised institutions, which may often be outpaced by the rate of (technological) change. Another challenge lies in centralised institutions' brittleness after they are established, that is, their vulnerability to regulatory capture or failure to react to changes in the issue area or technology. These issues are well reflected in challenges encountered in arms control regimes.

Establishing new international institutions is often a slow process, especially with higher participation and stakes. Under the General Agreement on Tariffs and Trade (GATT), negotiations for a 26 per cent cut in tariffs between 19 countries took 8 months in 1947. The Uruguay round, beginning in 1986, took 91 months to achieve a tariff reduction of 38 per cent between 125 parties (Martin and Messerlin, 2007). Historically, international law has been quicker at responding to technological change than to other changes; but even there its record is chequered, in some cases (e.g., spaceflight) adjusting within years, while being far more delayed in others (e.g., modern anti-personnel landmines) (Picker, 2001). Decentralised efforts might prove quicker to respond, especially if they rely more on informal institutions with a smaller, like-minded membership (Morin *et al*., 2019). Centralised governance may be particularly vulnerable to lengthy negotiations, especially if a few states hold unequal stakes in a technology, or if there are significant differences in information and expertise among state and private actors (Picker, 2001). AI fulfils both of these conditions. Moreover, because AI technology develops rapidly, slow implementation of rules and principles could enable certain actors to take advantage by setting de facto rules.

Even after its creation, a centralised regime can be brittle. The very qualities that provide it with political power may exacerbate the adverse effects of regulatory capture, and features that ensure institutional stability may also lead to an inability to adapt to new conditions. The regime might break before it bends. The first potential risk is regulatory capture. As illustrated by numerous cases, including undue corporate influence in the World Health Organisation during the 2009 H1N1 pandemic (Deshman, 2011), no institution is fully immune to capture, and centralisation may facilitate this by providing a single locus of influence (Martens, 2017). On the other hand, a regime complex comprising many smaller, parallel institutions could find itself vulnerable to capture by powerful actors, who can afford representation in every forum. Some have already expressed concern about the resources and sway of private tech actors in AI governance (Nemitz, 2018), and proposals for AI governance have been surrounded by calls to ensure their independence from such influence (Nature Editors, 2019).

Moreover, centralised regimes entail higher stakes. International institutions can be notoriously path-dependent and fail to adjust to changing circumstances (Baccaro and Mele, 2012). The public failure of a flagship global AI institution could have lasting political repercussions. It could strangle subsequent proposals in the crib, by undermining confidence in multilateral governance generally or on AI issues specifically. By contrast, for a decentralised regime complex to similarly fail, all of its component institutions would need to 'break' or fail to innovate simultaneously. A centralised institution that does not outright collapse, but which remains ineffective, may inhibit better efforts.

Ultimately, brittleness is not an inherent weakness of centralisation, but rather may depend on institutional design. There may be strategies to 'innovation-proof' (Maas, 2019a) governance regimes. Periodic renegotiation, modular expansion, additional protocols to framework conventions, 'principles based regulation', or sunset clauses can also support ongoing adaptation (see Marchant *et al*., 2011).

This discussion intersects with debates over whether a new centralised regime is even possible in today's shifting, dense institutional landscape (Alter and Raustiala, 2018; Morin *et al*., 2019). The speed of capability development in AI also highlights questions over the relative 'speed' or 'responsiveness' of different regime configurations. In slow-moving areas, a centralised regime's slowness may not be a problem. However, technological change has often 'perforated' many arms control regimes, from the Nuclear Non-Proliferation Treaty to the Missile Technology Control Regime, which sometimes struggled to carry out much-needed 'modernisation' in provisions or export control lists (Nelson, 2019).

This raises questions of necessary institutional speed. Is AI an issue that is so fast it makes centralisation untenable, such that we need a decentralised regime to match its speed and complexity? Or, should we use a singular institutional anchor to slow and channel the technology's development or application? There is precedent for international instruments directing or curtailing the development of certain technologies. The 1978 Environmental Modification Convention (ENMOD) Convention was an effective tool in preventing both funding for geoengineering research and the weaponised deployment of weather manipulation. By 1979, US investments in such technologies had dramatically decreased (Fleming, 2006).

The breadth vs. depth dilemma

Pursuing centralisation may create an overly high threshold that limits participation. Many multilateral agreements face a trade-off between having higher participation ('breadth') or stricter rules and greater ambition of commitments ('depth').

The dilemma is particularly evident for centralised institutions that are intended to be powerful and require strong commitments from states.

Sacrificing depth for breadth can also pose risks. The 2015 Paris Agreement on Climate Change was watered down to allow for the legal participation of the US. Anticipated difficulties in ratification through the Senate led to negotiators opting for a 'pledge and review' structure with few legal obligations, which permitted the US to join through executive approval (Kemp, 2017). In this case, inclusion of the US – which proved temporary – came at the cost of cutbacks to the demands which the regime made on all parties.

In contrast, decentralisation could allow for major powers to engage in at least some regulatory efforts, where they would be deterred from signing up to a more comprehensive package. This has precedence in climate governance. Some claim that the US-led Asia-Pacific Partnership on Clean Development and Climate helped, rather than hindered climate governance, as it bypassed the UN Framework Convention on Climate Change (UNFCCC) deadlock and secured (non-binding) commitments from actors not bound by the Kyoto Protocol (Zelli, 2011).

This matters, as buy-in may prove a particular thorny issue for AI governance. The actors who lead in AI development include powerful states, such as the US and China, that are potentially most adverse to restrictive global rules. They have thus far proved unenthusiastic regarding the global governance of security issues such as anti-personnel mines, LAWS, and cyberwarfare. In response, governance could take a different approach to military uses of AI. Rather than seeking a comprehensive agreement, devolving and spinning off certain components into separate treaties (e.g., separately covering LAWS testing standards; measures for liability and responsibility; or limits to operational context) could instead allow for the powerful to ratify and move forward some of those options (Weaver, 2014).

The breadth vs. depth dilemma is a trade-off in multilateralism generally, and a key challenge for centralisation. The benefit of a centralised body would be to create a powerful anchor that ensures policy coordination and coherence. In many cases, it will likely need to restrict membership to have teeth, or lose its teeth to secure wide participation. For specific issues in AI governance, this 'breadth vs. depth' trade-off might inform relative expectations of ongoing AI governance initiatives. If 'breadth' is more important, one might put more stock in nascent efforts at the UN (Garcia, 2020); if 'depth' of commitment seems more important, one might instead favour initiatives of like-minded states such as the GPAI.

The evolving architecture of AI governance suggests that a 'critical mass governance' (Kemp, 2017) approach may be appropriate. That is, there is a single centralised, framework under which progressive clubs move forward on particular issues. Rather than having an array of treaties, one has a set of protocols for different technologies or applications under a single framework. A similar approach has been taken in treaties such as the 1983 Convention on Long-Range Transboundary Air Pollution.

Forum shopping

Forum shopping may help or hinder AI governance. Fragmentation enables actors to choose where and how to engage. Such 'forum shopping' may take one of several forms: shifting venues, abandoning one, creating new venues, and working to sew competition among multiple (Braithwaite and Drahos, 2000). Even when there is a natural venue for an issue, actors have reasons to forum shop. For instance, states may look to maximise their influence (Pekkanen *et al*., 2007), and placate constituents by shifting to a toothless forum (Helfer, 2004). Membership in AI initiatives is highly varied and as initiatives begin to consider binding instruments, this ranging membership may be exploited.

The ability to successfully forum-shop depends on an actor's power. Most successful examples of forum-shifting have been led by the US (Braithwaite and Drahos, 2000). Intellectual property rights (IPR) in trade, for example, were subject to prolonged, contentious forum shopping. Developed states resisted attempts of the UN Conference on Trade and Development (UNCTAD) to address the issue by trying to shift it onto the World Intellectual Property Organisation (WIPO) (Braithwaite and Drahos, 2000) and then subsequently to the WTO (Helfer, 2004), despite protests from developing states. But weak states and non-state actors can also pursue forum shopping strategies in order to challenge the status quo, sometimes with success (Jupille *et al*., 2013). For example, developing states further shifted some IPR in trade to the WHO, and subsequently won concessions at the WTO (Kuyper, 2014).

Forum shopping may help or hurt governance (Gómez-Mera, 2016). This is evident in current efforts to regulate LAWS. While the Group of Governmental Experts has made some progress, on the whole the CCW has been slow. In response, activists have threatened to shift to another forum, as happened with the Ottawa Treaty that banned anti-personnel mines (Delcker, 2019). This strategy could catalyse progress, but also brings risks of further forum shopping. Forum shopping may similarly delay, stall, or weaken regulation of time-sensitive AI policy issues, including potential HLMI development. Non-state actors that participate in multiple fora may influence regime complex evolution, though perhaps to the detriment of other weak actors (Orsini, 2013). Thus, leading AI firms likely have sway when they elect to participate in some venues but not others. To date, leading AI firms appear to be prioritising engagement at the OECD over the UN. A decentralised regime will enable forum shopping, though further work is needed to determine whether this will help or hurt governance outcomes.

Policy coordination

There are good reasons to believe that either centralisation or fragmentation could enhance coordination. A centralised regime can enable easier coordination both across and within policy issues, acting as a focal point for states. Alternatively, fragmented institutions may be mutually supportive and even more creative.

Centralisation reduces the incidence of conflicting mandates and enables communication. These are the

ingredients for policy coherence, as shown in the case of the WTO above under 'political power'.

However, fragmented regimes can often act as complex adaptive systems. Political requests and communication between secretariats can ensure bottom-up coordination. Multiple organisations have sought to reduce greenhouse gas emissions within their respective remits, often at the behest of the UNFCCC Conference of Parties. Sometimes effective, bottom-up coordination can slowly evolve into centralisation. Indeed, this was the case for the GATT and numerous regional, bilateral and sectoral trade treaties, which all coalesced together into the WTO. While this organic self-organisation has occurred, it has taken decades.

Some have argued that 'polycentric' governance approaches may be more creative and legitimate than centrally coordinated regimes (Acharya, 2016; Ostrom, 2010). Arguments in favour of polycentricity include the notion that it enables governance initiatives to begin having impacts at diverse scales, and that it enables experimentation with policies and approaches (Ostrom, 2010). Consequently, these scholars assume 'that the invisible hand of a market of institutions leads to a better distribution of functions and effects' (Zelli and van Asselt, 2013, p. 7).

Yet an absence of centralised authority to manage regime complexes has presented challenges in the past. Across the proliferation of Multilateral Environmental Agreements (MEAs) there is no requirement to cede responsibility to the UN Environmental Programme in the case of overlap or competition. This has led to turf wars, inefficiencies and even contradictory policies (Biermann *et al*., 2009). One of the most notable examples is that of hydrofluorocarbons (HFCs). HFCs are potent greenhouse gases, and yet their use was encouraged by the Montreal Protocol since 1987 as a replacement for ozone-depleting substances. This was only resolved via the 2016 Kigali Amendment to the Protocol.

It is unclear if the different bodies covering AI issues will self-organise or collide. Many of the issues are interdependent and need to be addressed in tandem. Some policy-levers, such as regulating computing power or data, will impact multiple areas, given that AI development and use is closely tied to such inputs. Numerous initiatives on AI and robotics are displaying loose coordination (Kunz and Ó hÉigeartaigh, 2020). But it remains uncertain whether the virtues of a free market of governance will prevail. Great powers can exercise monopsony-like influence through forum shopping, and the supply of both computing power and machine learning expertise are highly concentrated. In sum, centralisation can reduce competition and enhance coordination, but it may suffocate the creative self-organisation of decentralised arrangements.

Discussion: what would history suggest?

Summary of considerations

The multilateral track record and peculiarities of AI yield suggestions and warnings for the future. A centralised regime could lower costs, support participation, and act as a powerful new linchpin within the international system. Yet centralisation could simply produce a brittle dinosaur, of symbolic value but with little meaningful impact. A poorly executed attempt at centralisation could lock-in a fate worse than fragmentation. Policy making and research alike could benefit from addressing the considerations presented in this paper, a summary of which is presented in Table 2.

The limitations of 'centralisation vs. decentralisation' debates

Structure is not a panacea. Specific provisions such as agendas and decision-making procedures matter greatly, as do the surrounding politics. Underlying political will may be impacted by framing or connecting policy issues (Koremenos *et al*., 2001b). The success of a regime depends on design details.

Moreover, institutions can be dynamic, and broaden over time by taking in new members or deepen in strengthening commitments. Successful multilateral efforts, such as trade and ozone depletion, tend to do both. Yet, decisions taken early on constrain and partially determine future paths. This dependency can even take place across regimes. The Kyoto Protocol was largely shaped by the targets-and-timetables approach of the Montreal Protocol, which itself drew from the Convention on Long-range Transboundary Air Pollution. This targets-and-timetables approach continues today in the way that most countries frame their climate pledges to the Paris Agreement. The choices we make on governing short-term AI challenges will likely shape the management of other policy issues in the long term (Cave and Ó hÉigeartaigh, 2019).

Yet, committing to centralisation, even if successful, may not solve the right problem – which may be geopolitical, not architectural. Centralisation could even exacerbate the problem by diluting scarce political attention, incurring heavy transaction costs, and shifting discussions away from bodies which have accumulated experience (Juma, 2000). For example, the Bretton Woods Institutions of the IMF and World Bank, joined later by the WTO, are centralised regimes that engender power. However, those institutions had the express support of the US and may have simply manifested state power in institutional form. Efforts to ban LAWS and create a cyberwarfare convention have been broadly opposed by states with an established technological superiority in these areas (Eilstrup-Sangiovanni, 2018).

HLMI: An illustrative example

The promise of centralisation may differ by policy issue. HLMI is one issue that is markedly unique: it is distinct in its risk profile, uncertainty, and linkage to other AI policy issues. While timelines are uncertain, the creation of such advanced AI systems is the express goal of various present-day projects (Baum, 2017), and the future development of an 'unaligned' HLMI could have catastrophic consequences (GCF, 2018). The creation of HLMI could lead to grotesque power imbalances. It could also exacerbate other AI policy

Table 2. Summary of considerations

Consideration	Implications for centralisation	Historical example	AI policy issue example
Political power	Pro	*Shaping other regimes:* WTO has created a chilling effect such that environmental treaties avoid trade-related measures.	Influencing powerful vested economic and military interests in AI may require a single empowered institution.
Efficiency & participation	Pro	*Decentralisation raises inefficiencies and barriers:* Proliferation of multilateral environmental agreements poses challenges in negotiation, implementation, and monitoring.	Fragmentation requires duplicative investment in AI subject-matter experts and undermines participation from developing countries and non-state actors.
Slowness & brittleness	Con	*Slowness:* Under the GATT, 1947 tariff negotiations among 19 countries took 8 months. The Uruguay round, beginning in 1986, took 91 months for 125 parties to agree on reductions. *Regulatory capture:* WHO accused of undue corporate influence in response to 2009 H1N1 pandemic.	Process of centralised regime development may not keep pace with the speed of AI development.
Breadth vs. depth dilemma	Con	*Watering down:* 2015 Paris Agreement suggests attempts to 'get all parties on board' may require less-stringent rules.	Attempts to effectively govern the military uses of AI have been resisted by the most powerful states.
Forum shopping	Depends on design	*Power predicts outcomes:* Developed countries shifted IPR in trade from UNCTAD to WIPO to WTO. *Accelerates progress:* NGOs and some states shifted away from CCW to ban anti-personnel mines.	Actors can use forum shopping to either undermine or catalyse progress on governance regimes for military AI systems.
Policy coordination	Depends on design	*Strong, but delayed convergence:* GATT and numerous trade treaties coalesced into the WTO after decades *Contradictory policies:* Montreal Protocol promoted the use of potent greenhouse gases for nearly thirty years.	Numerous AI governance initiatives display loose coordination, but it is unclear if these initiatives can respond to developments in a timely manner.

problems, such as labour automation and advanced military applications.

In Table 3 we provide a brief application of our framework to HLMI. It shows that centralisation of governance is particularly promising for HLMI. This is due to its neglect, stakes, scope, and need for informed, anticipatory policy.

Rather than any AI governance blueprint, our trade-offs framework provides one way of thinking through the costs and benefits of centralising governance. Identifying areas which are more easily defined and garner the benefits of centralised regulation provides an organic approach to thinking through which subset of topics an AI umbrella body could cover.

Lessons for theory

This is the first application of regime complex theory to the problem of AI governance. It is timely and pertinent given the nascent state of AI governance and of the technology itself. While the majority of the literature has observed mature regimes retrospectively, AI offers an opportunity for scholars to both track and influence the development of a new regime complex from its earliest stages.

Our analysis highlights both the uses and limits of the theoretical regime complex lens for AI. It can elucidate many important trade-offs, but provides little help in navigating the underlying geopolitics. The six considerations we have identified are also certainly not exhaustive of regime complex theory; further work could explore the complementary dynamics such as issue linkage, regime 'interplay management', or norm cascades in AI governance. Beyond this, the literature needs a better understanding of three key areas that are central to AI.

First, what does the political economy of AI mean for AI governance and centralisation? Regulatory capture is a genuine threat, yet many non-state actors hold valuable technical knowledge. Some, such as machine learning developers and NGOs have been influential in shaping governance on lethal autonomous weapons (Belfield, 2020). How these actors can shape the choice of fora and influence states under centralisation or decentralisation is pivotal.

Second, how should institutions match the speed of evolving collective action problems? Is the aim to make governance agile enough to keep pace with accelerating technological change or to manage the pace or direction of

Table 3. An application of the framework to high-level machine intelligence (HLMI)

Consideration	HLMI
Political power	Potential catastrophic risks make the increased political power of a centralised institution desirable. The creation of HLMI is a potential 'free-driver' issue. An effective response needs to have the teeth to deter major players from acting unilaterally. This will require a coordinated effort to track and forecast HLMI project efforts (see Baum, 2017), as well as a politically empowered organisation to act upon this information.
Efficiency & participation	Centralisation would support economies of scale in expertise to support efficient governance. Given the significant resources and infrastructure likely needed, a joint global development effort could be an efficient way to govern HLMI research.
Slowness & brittleness	If short HLMI timelines (less than 10-15 years) are expected, the lengthy period to negotiate and create such a body would be a critical weakness. If longer timelines are expected, there should be sufficient time to develop a centralised institution. Institutional capture is a concern given the resourced corporate actors involved in creating HLMI, e.g., Google or OpenAI. However, it is unclear if capture would be more likely under a centralised body.
Depth vs. breadth dilemma	Costs and requisite capabilities may restrict the development of HLMI to a few powerful players. Fewer actors makes centralisation more feasible. The breadth vs. depth dilemma could be avoided through a 'critical mass' approach that initially involves only the few countries that are capable of developing HLMI, although there would be legitimacy benefits to expanding membership.
Forum shopping	A centralised body is well placed to prevent forum shopping, as there is currently no coverage of HLMI development and deployment under international law. Future forum shopping could undermine timely negotiations amid risky HLMI development.
Policy coordination	Policy coordination is key for HLMI. It has close connections to issues such as labour automation and automated cyberwarfare. The creation or use of HLMI is not directly regulated by any treaties or legal instruments. This makes the creation of a new, dedicated institution to address it easier and less unlikely to trigger turf wars. However, it also makes it less likely that the existing tapestry of global governance can self-organise to cover HLMI in a timely manner.

such changes to levels that are socially and politically manageable? Theoretically, foresight methodologies have rarely been considered in regime complex debates. Yet for fast-moving and high-stakes technologies, they should be. Theory will need to better address how foresight and development trajectory monitoring capabilities intersect with the debates over governance architecture.

Third, how will these considerations look for particular institutional structures? We have presented a cursory case of HLMI and noted that there is an active debate of how to define AI and structure its governance. How will the case for centralisation look for a regime which targets just high-risk or military applications? Our framework provides an easily deployed way to analyse more discrete proposals for AI governance in the future.

Lessons for policy

Our framework provides a tool for policy makers to inform their decisions of whether to join, create, or forgo new AI policy institutions. For instance, the recent choice of whether to support the creation of an independent Global Panel on AI (GPAI) involved these considerations. Following the US veto at the G7 in 2019, GPAI was established in close relationship with the OECD. For now, it is worth monitoring the current landscape of AI governance to see if it exhibits enough policy coordination and political power to effectively deal with mounting AI policy problems. While there are promising initial signs (Kunz and Ó hÉigeartaigh, 2020) there are also already impending governance failures, such as for LAWS and cyberwarfare.

We outline a suggested monitoring method in Table 4. There are three areas to monitor: conflict, coordination, and catalyst. *Conflict* should measure the extent to which principles, rules, regulations, and other outcomes from different bodies in the AI regime complex undermine or contradict each other. *Coordination* seeks to measure the proactive steps that AI-related regimes take to work with each other. This includes liaison relationships, joint initiatives, and reinforcement between outputs and principles. *Catalyst* raises

Table 4. Regime complex monitoring suggestions

Key theme	Questions	Methods
Conflict	To what extent are regimes' principles and outputs in opposition over time?	Expert and practitioner survey Network analysis (e.g., citation network clustering and centrality)
Coordination	Are regimes taking steps to complement each other?	Natural Language Processing (e.g., textual entailment and fact checking)
Catalyst	Are regimes self-organizing to proactively fill governance gaps?	

the important question of governance gaps: is the regime complex self-organising to proactively address international AI policy problems? Numerous AI policy problems currently have no clear coverage under international law. Monitoring these regime complex developments, using various existing and emerging tools (see Maas, 2019b; Deeks, 2020), could inform a discussion and decision of whether to centralise AI governance further.

The international governance of AI is nascent and fragmented. Centralisation under a well-designed, modular, 'innovation-proof', critical mass framework may be a desirable solution. However, such a move must be approached with caution. Defining its scope and mandate is one problem. Ensuring a politically-acceptable and well-designed body is perhaps a more daunting one. For now, we should closely watch the trajectory of both AI technology and its governance initiatives to determine whether centralisation is worth the risk.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Notes

The authors thank Seth Baum, Haydn Belfield, Jessica Cussins-Newman, Martina Kunz, Jade Leung, Nicolas Moës, Robert de Neufville, Nicolas Zahn, and participants of the AAAI/ACM Conference on AI, Ethics and Society 2020 for valuable comments. They also thank two anonymous reviewers for excellent and detailed comments. Any remaining errors are the authors' alone. This publication was made possible through the support of a grant from Templeton World Charity Foundation, Inc. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation, Inc. No conflict of interest is identified.

1. A regime is a set of 'implicit or explicit principles, norms, rules and decision-making procedures around which actors' expectations converge in a given area of international relations' (Krasner, 1982, p. 186).
2. We define 'AI' as any machine system capable of functioning 'appropriately and with foresight in its environment' (Nilsson, 2009, p. 13).
3. 'High-level machine intelligence' has been defined as 'unaided machines [that] can accomplish every task better and more cheaply than human workers' (Grace *et al*., 2018, p. 731).
4. However, China has endorsed the G20 AI Principles, which reflect the OECD Principles.

References

Acharya, A. (2016) 'The Future of Global Governance: Fragmentation May be Inevitable and Creative', *Global Governance: A Review of Multilateralism and International Organizations*, 22 (4), pp. 453–460.

Alter, K. J. and Raustiala, K. (2018) 'The Rise of International Regime Complexity', *Annual Review of Law and Social Science*, 14 (1), pp. 329–349.

Anton, D. (2012) ''Treaty Congestion' in International Environmental Law', in S. Alam, J. H. Bhuiyan, T. M. Chowdhury and E. J. Techera (eds.), *Routledge Handbook of International Environmental Law*. Abingdon: Routledge, pp. 651–666.

Baccaro, L. and Mele, V. (2012) 'Pathology of Path Dependency? The ILO and the Challenge of New Governance', *ILR Review*, 65 (2), pp. 195–224.

Baum, S. (2017) 'A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy'. Global Catastrophic Risk Institute Working Paper, pp. 1–17. Available from: https://papers.ssrn.com/abstract=3070741 [Accessed 29 October 2020].

Belfield, H. (2020) 'Activism by the AI Community: Analysing Recent Achievements and Future Prospects'. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 15–21.

Biermann, F. and Kim, R. (2020) *Architectures of Earth System Governance. Institutional Complexity and Structural Transformation*. Cambridge: Cambridge University Press.

Biermann, F., Pattberg, P., van Asselt, H. & Zelli, F. (2009) 'The Fragmentation of Global Governance Architectures: A Framework for Analysis', *Global Environmental Politics*, 9 (4), pp. 14–40.

Braithwaite, J. and Drahos, P. (2000) *Global Business Regulation*. Cambridge: Cambridge University Press.

Brockmann, K. (2019) *Challenges to Multilateral Export Controls*. Stockholm: Stockholm International Peace Research Institute.

Calo, R. (2017) 'Artificial Intelligence Policy: A Primer and Roadmap', *UC Davis Law Review*, 51 (2), pp. 399–435.

Cave, S. and Ó hÉigeartaigh, S. (2019) 'Bridging Near- and Long-term Concerns About AI', *Nature Machine Intelligence*, 1 (1), p. 5–6.

Cihon, P. (2019) 'Standards for AI Governance: International Standards to Enable Global Coordination in AI Research and Development', Technical Report, Center for the Governance of AI, Future of Humanity Institute, Oxford. Available from: https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf [Accessed 29 October 2020].

Council of Europe (COE) (2020) 'Ad Hoc Committee on Artificial Intelligence'. Available from: https://rm.coe.int/leaflet-cahai-en-june-2020/16809eaf12 [Accessed 29 October 2020].

Deeks, A. (2020) 'High-Tech International Law', *George Washington Law Review*, 88 (3), pp. 574–653.

Delcker, J. (2019) 'How Killer Robots Overran the UN', *POLITICO*. Available from: https://www.politico.eu/article/killer-robots-overran-united-nations-lethal-autonomous-weapons-systems/ [Accessed: 29 October 2020].

Delcker, J. (2020) 'POLITICO AI: Decoded: Trump's White House Throws Its Weight behind OECD AI Efforts – Lobbying Battle over Europe's AI Rules – Fighting Hate Speech with AI', *POLITICO*. Available from: https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-trumps-white-house-throws-its-weight-behind-oecd-ai-efforts-lobbying-battle-over-europes-ai-rules-fighting-hate-speech-with-ai/ [Accessed: 29 October 2020].

Deshman, A. C. (2011) 'Horizontal Review between International Organizations: Why, How, and Who Cares about Corporate Regulatory Capture', *European Journal of International Law*, 22 (4), pp. 1089–1113.

Drezner, D. W. (2009) 'The Power and Peril of International Regime Complexity', *Perspectives on Politics*, 7 (1), pp. 65–70.

Eckersley, R. (2004) 'The Big Chill: The WTO and Multilateral Environmental Agreements', *Global Environmental Politics*, 4 (2), pp. 24–50.

Eilstrup-Sangiovanni, M. (2018) 'Why the World Needs an International Cyberwar Convention', *Philosophy & Technology*, 31 (3), pp. 379–407.

Ekstrom, J. A. and Crona, B. I. (2017) 'Institutional Misfit and Environmental Change: A Systems Approach to Address Ocean Acidification', *Science of the Total Environment*, 576, pp. 599–608. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0048969716323014?via%3Dihub [Accessed 02 November 2020].

Erdelyi, O. and Goldsmith, J. (2018) 'Regulating Artificial Intelligence: Proposal for a Global Solution', in *Proceedings of the 2018 AAAI / ACM Conference on Artificial Intelligence, Ethics and Society*', pp. 1–7. Available from: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf [Accessed 15 May 2018].

Esty, D. and Ivanova, M. (2002) 'Revitalizing Global Environmental Governance: A Function-Driven Approach', in D. C. Esty and M. H.

Ivanova (eds.), *Global Environmental Governance: Options & Opportunities*. New Haven, CT: Yale School of Forestry and Environmental Studies, pp. 181–204. Available from: https://environment.yale.edu/publication-series/documents/downloads/a-g/esty-ivanova.pdf [Accessed 23 September 2019].

Fleming, J. (2006) 'The Pathological History of Weather and Climate Modification: Three Cycles of Promise and Hype', *Historical Studies in the Physical and Biological Sciences*, 37 (1), pp. 3–25.

Garcia, E. (2020) 'Multilateralism and Artificial Intelligence: What Role for the United Nations?', in M. Tinnirello (ed.), *The Global Politics of Artificial Intelligence*. Boca Raton: CRC Press, pp. 1–20.

GCF (2018) 'Global Catastrophic Risks 2018', Technical report, Global Challenges Foundation. Available from: https://www.humanitarianfutures.org/wp-content/uploads/2018/11/GCF-Annual-report-2018.pdf [Accessed 11 December 2019].

Global Partnership on Artificial Intelligence (GPAI) (2020) *Joint Statement from Founding Members of the Global Partnership on Artificial Intelligence*. Press release. Available from: https://www.diplomatie.gouv.fr/en/french-foreign-policy/digital-diplomacy/news/article/launch-of-the-global-partnership-on-artificial-intelligence-by-15-founding [Accessed 15 June 2020].

Gómez-Mera, L. (2016) 'Regime Complexity and Global Governance: The Case of Trafficking in Persons', *European Journal of International Relations*, 22 (3), pp. 566–595.

Gómez-Mera, L., Morin, J. and Van De Graaf, T. (2020) 'Regime Complexes', in F. Biermann and R. E. Kim (eds.), *Architectures of Earth System Governance: Institutional Complexity and Structural Transformation*. Cambridge: Cambridge University Press, pp. 137–157.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B. and Evans, O. (2018) 'When will AI Exceed Human Performance? Evidence from AI Experts', *Journal of Artificial Intelligence Research*, 62, pp. 729–754.

Green, J. and Auld, G. (2017) 'Unbundling the Regime Complex: The Effects of Private Authority', *Transnational Environmental Law*, 6 (2), pp. 259–284.

Helfer, L. (2004) 'Regime Shifting: The TRIPs Agreement and New Dynamics of International Intellectual Property Lawmaking', *Yale Journal of International Law*, 29 (1), pp. 1–83.

ITU (2019) 'United Nations Activities on Artificial Intelligence (AI) 2019', Technical report, ITU. Available from: https://www.itu.int/dms/pub/itu-s/opb/gen/SGEN-UNACT-2019-1-PDF-E.pdf [Accessed 13 November 2019].

Jelinek, T., Wallach, W. and Kerimi, D. (2020) 'Policy brief: the creation of a G20 coordinating committee for the governance of artificial intelligence', AI and Ethics. Available from: https://link.springer.com/article/10.1007/s43681-020-00019-y#citeas [30 October 2020].

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence*, 1 (9), pp. 389–399.

Juma, C. (2000) 'Commentary: The Perils of Centralizing Global Environmental Governance', *Environment: Science and Policy for Sustainable Development*, 42 (9), pp. 44–45.

Jupille, J., Mattli, W. and Snidal, D. (2013) '*Institutional Choice and Global Commerce*', Cambridge: Cambridge University Press.

Kemp, L. (2017) 'US-proofing the Paris Climate Agreement', *Climate Policy*, 17 (1), pp. 86–101.

Kemp, L., Cihon, P., Maas, M.*et al*. (2019) 'UN High-level Panel on Digital Cooperation: A Proposal for International AI Governance'. Available from: https://digitalcooperation.org/wp-content/uploads/2019/02/LukeKempSubmission-to-the-UN-HighLevel-Panel-on-Digital-Cooperation-2019-Kemp-etal.pdf [Accessed 4 March 2019].

Kennan, G. (1970) 'To Prevent A World Wasteland: A Proposal', *Foreign Affairs*, 48 (401), pp. 409–412.

Koremenos, B., Lipson, C. and Snidal, D. (2001a) 'Rational Design: Looking Back to Move Forward', *International Organization*, 55 (4), pp. 1051–1082.

Koremenos, B., Lipson, C. and Snidal, D. (2001b) 'The Rational Design of International Institutions', *International Organization*, 55 (4), pp. 761–799.

Krasner, S. D. (1982) 'Structural Causes and Regime Consequences: Regimes as Intervening Variables', *International Organization*, 36 (2), pp. 185–205.

Kunz, M. and Ó hÉigeartaigh, S. (2020) 'Artificial Intelligence and Robotization', in R. Geiss and N. Melzer (eds.), *Oxford Handbook on the International Law of Global Security*. Oxford: Oxford University Press, pp. 1–16.

Kuyper, J. (2014) 'Global Democratization and International Regime Complexity', *European Journal of International Relations*, 20 (3), pp. 620–646.

Maas, M. (2019a) 'Innovation-proof Governance for Military AI? How I Learned to Stop Worrying and Love the Bot', *Journal of International Humanitarian Legal Studies*, 10 (1), pp. 129–157.

Maas, M. (2019b) 'International Law Does Not Compute: Artificial Intelligence and The Development, Displacement or Destruction of the Global Legal Order', *Melbourne Journal of International Law*, 20 (1), pp. 29–56.

Marchant, G., Allenby, B. and Herkert, J. (2011) *The Growing Gap Between Emerging Technologies and Legal-ethical Oversight: The Pacing Problem*, Vol. 7. New York: Springer Science & Business Media.

Martens, J. (2017) *Corporate Influence on the G20: The Case of the B20 and Transnational Business Networks*. Berlin: Heinrich-Böll-Stiftung and Global Policy Forum.

Martin, W. and Messerlin, P. (2007) 'Why is it so Difficult? Trade Liberalization Under the Doha Agenda', *Oxford Review of Economic Policy*, 23 (3), pp. 347–366.

Morin, J., Dobson, H., Peacock, C., Prys-Hansen, M., Anne, A., Bélanger, L. *et al*. (2019) 'How Informality Can Address Emerging Issues: Making the Most of the G7', *Global Policy*, 10 (2), pp. 267–273.

Nature Editors (2019) 'International AI Ethics Panel Must Be Independent', *Nature*, 572, p. 415. Available from: https://www.nature.com/articles/d41586-019-02491-x [Accessed 02 September, 2019].

Nelson, A. (2019) '*Innovation Acceleration, Digitization, and the Arms Control Imperative*'. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. Available from: https://papers.ssrn.com/abstract=3382956 [Accessed 26 March, 2019].

Nemitz, P. (2018) 'Constitutional Democracy and Technology in the Age of Artificial Intelligence'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376 (2133), p. 20180089.

Nilsson, N. (2009) *The Quest for Artificial Intelligence*. New York, NY: Cambridge University Press.

OECD (2020a) 'OECD Principles on Artificial Intelligence [online]'. Available from: https://www.oecd.org/going-digital/ai/principles/ [Accessed 17 August 2020].

OECD (2020b) 'OECD Network of Experts on AI [online]'. Available from: https://oecd.ai/network-of-experts [Accessed 17 August 2020].

Orsini, A. (2013) 'Multi-forum Non-state Actors: Navigating the Regime Complexes for Forestry and Genetic Resources', *Global Environmental Politics*, 13 (3), pp. 34–55.

Orsini, A., Morin, J. and Young, O. (2013) 'Regime Complexes: A Buzz, a Boom, or a Boost for Global Governance?', *Global Governance: A Review of Multilateralism and International Organizations*, 19 (1), pp. 27–39.

Ostrom, E. (2010) 'Polycentric Systems for Coping with Collective Action and Global Environmental Change', *Global Environmental Change*, 20 (4), pp. 550–557.

Pauwels, E. (2019) 'The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI', Technical report, United Nations University Centre for Policy Research. Available from: https://i.unu.edu/media/cpr.unu.edu/attachment/3472/PauwelsAIGeopolitics.pdf [Accessed 11 July 2019].

Pekkanen, S. M., Solıs, M. and Katada, S. N. (2007) 'Trading Gains for Control: International Trade Forums and Japanese Economic Diplomacy', *International Studies Quarterly*, 51 (4), pp. 945–970.

Picker, C. (2001) 'A View from 40,000 Feet: International Law and the Invisible Hand of Technology', *Cardozo Law Review*, 23, pp. 151–219.
Rayfuse, R. (2017) 'Public International Law and the Regulation of Emerging Technologies'. In *The Oxford Handbook of Law, Regulation and Technology*, 2017. Available from: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199680832.001.0001/oxfordhb-9780199680832-e-22 [Accessed 3 January 2019].
Schiff, D., Biddle, J., Borenstein, J. and Laas, K. (2020) 'What's Next for AI Ethics, Policy, and Governance? A Global Overview', in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 153–158.
Schuett, J. (2019) 'A Legal Definition of AI', ArXiv:1909.01095 [Cs]. Available from: http://arxiv.org/abs/1909.01095 [Accessed 6 January 2020].
Stilgoe, J., Owen, R. and Macnaghten, P. (2013) 'Developing a Framework for Responsible Innovation', *Research Policy*, 42 (9), pp. 1568–1580.
Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G. *et al* (2016) 'Artificial Intelligence and Life in 2030', Technical report, Stanford University. Available from: http://ai100.stanford.edu/2016-report [Accessed 26 February 2017].
UNESCO (2020) *Ad Hoc Expert Group (AHEG) for the Preparation of a Draft Text of a Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.
United Nations Secretary-General (2020) *Road Map for Digital Cooperation: Implementation of the Recommendations of the High-level Panel on Digital Cooperation*. New York, NY: United Nations.
Wallach, W. and Marchant, G. (2018) 'An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics', in *Proceedings of the 2018 AAAI / ACM Conference on Artificial Intelligence, Ethics and Society*, pp. 1–7. Available from: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_77.pdf [Accessed 14 June 2018].
Weaver, J. (2014) 'Autonomous Weapons and International Law: We Need These Three International Treaties to Govern 'Killer Robots'', *Slate Magazine*. Available from: https://slate.com/technology/2014/12/autonomous-weapons-and-international-law-we-needthese-three-treaties-to-govern-killer-robots.html [Accessed 16 April 2019].
Young, O. (2002) *The Institutional Dimensions of Environmental Change: Fit, Interplay, and Scale*. Cambridge, MA: The MIT Press.
Zelli, F. (2011) 'The Fragmentation of the Global Climate Governance Architecture', *Wiley Interdisciplinary Reviews: Climate Change*, 2 (2), pp. 255–270.
Zelli, F. and van Asselt, H. (2013) 'Introduction: The Institutional Fragmentation of Global Environmental Governance: Causes, Consequences, and Responses', *Global Environmental Politics*, 13 (3), pp. 1–13.
Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the Frontier of Power*. London: Profile Books.

Author Information

Peter Cihon is a Policy Analyst at GitHub. He wrote the paper as a Research Affiliate with the Centre for the Governance of AI housed at the Future of Humanity Institute, University of Oxford; it does not reflect the views of his current employer. He holds an M.Phil. in Technology Policy from the University of Cambridge and a B.A. in Sociology and Political Economy from Williams College.

Matthijs M. Maas is a Research Associate at the Centre for the Study of Existential Risk at the University of Cambridge, and wrote this paper as a PhD Fellow at the Centre for International Law and Governance, at the University of Copenhagen Faculty of Law, as well as a Research Affiliate with the Centre for the Governance of AI (Future of Humanity Institute, University of Oxford). He holds a Msc in International Relations from the University of Edinburgh, and a B.A. Liberal Arts from University College Utrecht. Matthijs focuses on strategies to ensure the efficacy, resilience and coherence of AI governance regimes.

Luke Kemp is a Research Associate at the Centre for the Study of Existential Risk at the University of Cambridge. He holds a Doctorate in International Relations and a Bachelor of Interdisciplinary Studies with first-class honours from the Australian National University. Luke focuses on societal collapse and foresight for catastrophic risks.