Opinion – A New International AI Body Is No Panacea

https://www.e-ir.info/2023/08/11/opinion-a-new-international-ai-body-is-no-panacea/

HUW ROBERTS, AUG 11 2023

Late 2022 and early 2023 saw breakthroughs in the commercialisation of foundation models: a new type of artificial intelligence (AI) system designed to be adaptable for a wide range of downstream tasks. Foundation models have underpinned the development of new products, like OpenAI's ChatGPT, while also being integrated into the existing products of a variety of companies, including Microsoft's search engine Bing. These systems enhance consumer experiences and improve business efficiency, yet they also bring about new risks. Foundation models democratise capabilities that can be used to conduct advanced cyberattacks and produce disinformation, they have the potential to reinforce harmful biases and displace jobs in a way that exacerbates national and international inequalities, and could hasten climate change due to the environmental footprint of developing these systems.

The risks associated with this new form of AI has reenergised calls for a new international AI body. Prominent figures, including the United Nations (UN) Secretary General Antonio Guterres, OpenAI's CEO Sam Altman, and British Prime Minister Rishi Sunak, have all argued for the creation of a new international AI body modelled on institutions like the International Atomic Energy Agency (IAEA). This body, they variously claim, has the potential to mitigate risks ranging from harmful bias to existential threats to humanity. High-profile attention on international AI cooperation is a positive step forward, but a new AI body is no panacea. In fact, excessive attention on establishing a new AI body may distract from other types of institutional reform that could more viably support positive outcomes from AI.

International institutions are a product of their time. In the years following the Second World War, breakthroughs in technology, combined with a post-war impetus, led to globalised connectivity and an associated demand for institutions. The IAEA was established in 1957, during this period of proliferation in international institutions, to manage the global opportunities and risks of nuclear technologies. Nowadays, formal institution making is rare. Decolonialisation, institutional inertia, fragmentation, and the increased complexity of problems that result from modern connectivity have all contributed to a "gridlock" which hinders the kind of formal institution-making that was seen in aftermath of the Second World War. States are now more reliant on informal multilateral agreements (i.e., not treaty-based) and nongovernmental institutions to promote international cooperation. Any effort to create an IAEA-like body with authoritative standards-making and monitoring powers would likely require the type of formal interstate agreement that is now infrequently seen.

The characteristics of AI as a policy area further complicate the possibility of establishing a new and meaningfully empowered international AI body. There are four key reasons for this. First, AI is a general-purpose technology that impacts all aspects of society, leading states to perceive it as being central to international competition. China has singled out AI as a technology that can facilitate the "leapfrogging" of competitors, with a policy of military-civil fusion designed to ensure that breakthroughs in civilian technologies can support military applications. The US has responded by introducing policies designed to hamstring China's AI development, particularly through export restrictions on semiconductors. This type of competitive framing and policymaking undermines trust and the potential for cooperation, as was implicit in comments by China's delegate at the first UN Security Council meeting on AI.

Second, while AI presents serious international risks, there is currently no shared perception of it posing an existential threat. This means that there is a lesser incentive for interstate cooperation when compared to other historic cases of cooperation over strategic technologies, notably, arms control for nuclear weapons during the Cold

War. During this period, the tangible threat of global destruction helped progress discussions and policy action.

Third, AI governance is a relatively new policy area that has shallow institutional foundations. Institutional change often takes place incrementally through a process of gradual "layering" on top of what already exists rather than through a rapid "punctuated equilibrium" (van der Heijden, 2011). Take the Intergovernmental Panel on Climate Change (IPCC) – another model for an international AI body touted by scholars – which was preceded by almost two decades of multilateral scientific assessments, before being formalised into the IPCC. Given the recency of multilateral AI governance work, there is currently no initiative that could be easily formalised to become a comparable international body.

Fourth, AI is not a single policy problem, but rather a set of problems. Good AI governance involves mitigating harmful biases, establishing rules for lethal autonomous weapons, checking anti-competitive behaviours from technology companies, and many other issues. It would be extremely ambitious to develop an institution capable of tackling all these problems, with examples of centralised global governance in other policy areas, like trade, emerging out of a long process of layering. Indeed, it is arguable whether a centralised model for international governance is the most effective method for dealing with such an array of policy problems, as it may suffer from brittleness when new issues arise. This indicates that even if a new AI body could be established, it would provide a partial solution at best.

Given the difficulties associated with establishing a new and meaningfully empowered international AI governance body, discussion should move away from what an idealised institution could look like, towards how existing initiatives can realistically be built upon to bring about positive change. Early work by the OECD, G20, and UNESCO, among others, is beginning to provide a foundation for global AI governance. However, initiatives have thus far been high-level and fragmented, limiting their effectiveness. For example, both the OECD and UNESCO have produced guiding principles designed to ensure that the development and use of AI is ethical, signalling a lack of authoritative guidance and a waste of resource due to duplicate efforts.

A focus on mitigating fragmentation through strengthening coordination between institutions and coalescing around common goals has the potential to bring mutually reinforcing change that is more than a sum of its parts. In this decentralised but coordinated model, international organisations develop authoritative policy guidance and scrutiny mechanisms related to their remits, which provide clarity and puts pressure on state, subnational, and transnational stakeholders to enact unilateral initiatives. These individual efforts coalesce around common goals to produce incremental but transformative change.

While a decentralised approach may sound ineffective, there is a strong precedent of it being used to circumvent gridlock in adjacent policy areas, notably, climate governance. After failed attempts at reaching binding global agreement on cutting emissions, the 2015 Paris Accords cemented a shift towards the type of "catalytic model" described above. This decentralised model has allowed for progress to be made in some climate policy areas, even while political conditions stall progress in others. Emulating a similar strategy for international AI governance is an imperfect solution for the challenges faced, but under current geopolitical and institutional conditions, it is the most viable path forward for managing the opportunities and risks of AI.

The UN's new Multistakeholder Advisory Body on AI provides a valuable outlet for progressing discussions on strengthening international coordination; for instance, through mapping which institutions are currently fulfilling international AI governance functions and providing recommendations for how gaps can be filled and duplication lessened. It is imperative that this new body is used to push for realistic change rather than promoting pipedreams.

About the author:

Huw Roberts is a doctoral researcher at the University of Oxford's Internet Institute. His research focuses on

comparative and international AI governance. Huw previously worked for the UK Government where he co-wrote several AI policy documents, including the country's National AI Strategy.