



POLICY BRIEF **MANAGING THE TRANSITION TO A MULTI-STAKEHOLDER ARTIFICIAL INTELLIGENCE GOVERNANCE**



Task Force 5
**THE FUTURE OF MULTILATERALISM AND
GLOBAL GOVERNANCE**

Authors

MARÍA BELÉN ABDALA, ANDRÉS ORTEGA, JULIA POMARES

موجز السياسة إدارة التحول نحو حوكمة ذكاء اصطناعي لأصحاب المصلحة المتعددين

فريق العمل الخامس
مستقبل التعددية والحوكمة العالمية



المؤلفون
ماريا بيلين أبدال، أندريس أورتيجا، جوليا بوماريس



ABSTRACT

Artificial intelligence (AI) must be governable and interoperable to ensure that it reduces existing inequalities without creating new divides. Building on the 2019 Group of Twenty (G20) Communiqué and the Organization for Economic Cooperation and Development (OECD) guidelines, this policy brief presents a roadmap for the practical implementation of AI regulation.

A multi-stakeholder approach towards AI governance can contribute to establishing legitimate and trusted global standards that function as decisive tools in stewarding increasingly digitally oriented societies towards complete social inclusion. The challenge is transnational and thus, cooperation must also flow across borders. The G20 has a key role in this endeavor.

يجب أن يكون الذكاء الاصطناعي قابلاً للحكومة وقابلًا للتشغيل المتبادل لضمان قدرته على الحد من حالات عدم المساواة القائمة دون إحداث فجوات جديدة. وانطلاقاً من البيان الرسمي لمجموعة العشرين لعام ٢٠١٩ والمبادئ التوجيهية لمنظمة التعاون الاقتصادي والتنمية (OECD)، يعرض موجز السياسة هذا خارطة طريق للتطبيق العملي للوائح التنظيمية للذكاء الاصطناعي.

يمكن أن يساهم نهج أصحاب المصلحة المتعددين بخصوص حوكمة الذكاء الاصطناعي في وضع معايير عالمية شرعية وموثوق بها، والتي تعمل كأدوات حاسمة في توجيه المجتمعات ذات التوجه الرقمي نحو الإدماج الاجتماعي الكامل. وهذا التحدي عابر للحدود، وبالتالي، يجب أن يتخطى التعاون الحدود أيضاً. حيث تؤدي مجموعة العشرين دوراً رئيسياً في هذا المسعى.



CHALLENGE

Artificial intelligence (AI) refers to the designing and building of intelligent agents that receive precepts from the environment and take action in response to the detected contexts (Russell and Norvig 2009). While AI is hardly a recent invention, it is being increasingly used in the provision of public services to support decision-making processes, to interact with citizens, or to streamline government procedures. However, most systems that were developed are characterized by low levels of transparency, public awareness, and supervision and liability measures. The use of complex and opaque algorithms is an important challenge in “normal” times. It becomes even more important given the role of AI in many countries as part of the fight against the coronavirus disease (COVID-19) pandemic through special applications (apps) and other means of data selection.

The future depends not only on technical advancements, but also on collective intelligence and political choices (Mulgan 2018). The successful incorporation of AI requires public administrations to redefine strategies based on the use of new technologies and to develop adequate governance structures. The Group of Twenty (G20) leaders have already moved in accordance with this motivation. They have expressed the need to help societies adapt to the digital transformation of our economies in the 2019 Communiqué and endorsed the OECD’s Artificial Intelligence Principles, stating the need for an AI centered on people.

Despite these efforts, little progress has been made on the governance aspect of AI and the regulations that are needed to reach those objectives. Moreover, differences across the G20 countries are already surfacing in terms of their capacity to design and use AI, and the elaboration of principles and strategies for its development at the national, regional, and local levels.¹ The lack of coordination can lead to a fragmented governance landscape that exacerbates pre-existing inequities, prevents citizens from accessing equal rights across jurisdictions, and produces new types of divides

1. Estonia, for instance, brought together a group of experts from the public and private sectors to work on a bill that encompasses AI in a comprehensive manner. Singapore recognized the need for a regulatory framework for AI, but initially adopted a lighter approach meant to promote its further development. China presented the objectives of its plan and launched a code of conduct addressing the questions of the values of AI. The state of California, in the US, enacted one of the strictest laws on personal data protection, emulating the European General Data Protection Regulation (Crawford et al. 2019). For a detailed description, see Pomares and Abdala (2020).

CHALLENGE

between countries and regions (Cihon 2018). Likewise, a race in which governments compete to attract AI industries through national incentives may fail to foster the oversight needed to mitigate the risks associated with these processes (Bostrom 2014). It is not only a technical issue, but a cultural and geopolitical one.

AI must be governable and, therefore, defining common global standards and policy options that foster dialogue with both public and private institutions is decisive in stewarding increasingly digitally oriented societies toward complete social inclusion.



PROPOSAL

The discussion about the way ahead involves regulatory and policy options, which can be interpreted as the “sustained and focused control exercised by a public agency over activities that are valued by the community” (Selznick 1985). The goal is to ensure the sustainable, transparent, and inclusive development of AI in governments, while reducing existing inequalities rather than creating new divides. The policy options discuss four important aspects: 1) identifying the key concepts and systems subject to one or more of the procedures, 2) the exercise of control, 3) the agents responsible for this control, and 4) the integrity of the process itself.

Proposal I

Establish a common language and define the key concepts surrounding AI

Governments are recognizing the need for the governance that AI creates, but the national and regional initiatives produced so far lead to rather abstract statements based on principles and values, possessing few specific recommendations. Ideas such as “fairness” and “security” are contested concepts that require contextual interpretation. Agreements on their relevance hide the sources of political and ethical conflict contained in each key concept (Imbrie and Kania 2019; Mittelstadt 2019) and conceal a geopolitical race as values generated in a sociopolitical context fuel soft as well as hard military power (Ortega 2020). This ambiguity will make it difficult to translate vague principles into concrete action.

Therefore, to define the subject of the regulation, a first step would be to develop a common understanding of the meaning of the core concepts for AI development and governance. The G20 functions as a forum for executive debate that encompasses the major economies and more relevant political powers of the world. It could convene technical experts from academia and industry to explore shared concepts, concerns, and research agendas that involve key issues on AI systems. It could consider alternate framings and build the appropriate terminology, while also identifying the issues in dispute. Such a multi-stakeholder and multilevel engagement can facilitate future dialogue and could become the basis for further cooperation in which AI serves collective intelligence, not the other way around. Over time, the presence of clear, shared definitions of the factors that influence the design, development, and the deployment of AI techniques can help improve transparency and promote a foundation for continued and collaborative initiatives to promote AI safety and security. This cooperation would also allow the G20 to become a reservoir of AI knowledge. This can be achieved by sharing the best practices to guide technical innovation among and within G20 countries, developing policy approaches to deal with common concerns, and monitoring its implementation.

Proposal II

Move beyond ethical principles

A second aspect refers to the exercise of control, and the tools through which this can be achieved. AI governance is not only a matter of setting up common general ethical principles, but also of operationalizing them and embedding them in the programs and in the real algorithms.

In addition to the conceptual differences already mentioned, the operationalization of these principles is a complex process in its early stages. Their implementation is based on the assumption of goodwill from the actors involved and does not propose concrete incentives to achieve the goals that are being set. The gap between intent and practice is large and its documentation remains lacking (Fjeld and Nagy 2020; Mittelstadt 2019). The proposals advocated by the European Commission (2019) and the Institute of Electrical and Electronic Engineers (IEEE 2019) are more advanced, in particular, in the difficult issue of translating principles into actual code and algorithms.

To move beyond principles and fully address these challenges, a certain degree of international agreement is required. Common rules and regulations could potentially take the form of standards (Austin 1995; Baldwin, Cave, and Lodge 2012) that help define technical systems and usher in their social impact (Cihon 2018). Standards can provide the guidelines required to develop new technologies as well as the safety procedures to foster this in a controlled manner. By establishing shared ground-rules, they can reduce the risks in both international and market competition, thereby supporting policy goals where global governance (the governance of increasingly automated decision systems) is needed. International standards could change the context in which AI is researched, developed, and implemented, simultaneously dealing with the geopolitical and cultural differences at hand and disseminating best practices at the global level.

The scope for standards is not predetermined, and many forms can impact the development of automated systems such as ethical, humanitarian, legal, and political normative frameworks. They do not have to be completely uniform globally. The important concept is “interoperability,” which denotes a common ground from which AI algorithms and machines can operate together.

As the G20 leaders stated in the 2019 G20 Ministerial Statement on Trade and Digital Economy, “governance in the digital era needs to be not only innovation-friendly but

also innovative itself, while not losing legal certainty.” International standards, frameworks, and regulatory cooperation can help in this regard. Especially in the current context, in which key governments, including China and the US,² have voiced priorities for developing international AI standards which may become increasingly contentious over time, as has been witnessed in telecommunications.

Proposal III

Map the agents and organizations responsible for the governance of AI

The process of generating standards has often been conceived as a technical one. However, defining standards often carry significant implications on the trajectories on which technologies and markets evolve, providing leverage to those who lead the development of standardized technologies (Seaman 2020). Thus, the ability to define standards is a part of international power competition.³ As the field is therefore susceptible to strategic and geopolitical considerations, generating a multi-stakeholder approach for the creation, dissemination, and enforcement of standards can foster trust among states, researchers, and possible competitors.

Specialized agencies can be created to carry the task forward. However, there are already international standards bodies in place who govern socio-technical issues and possess the institutional capacity to achieve expert consensus and both propagate and enforce standards across the world. Examples of these are the International Organization for Standardization, an independent NGO, and the International Electrotechnical Commission. Moreover, the IEEE Standards Association (an engineers’ professional organization that has addressed protocols for products, software engineering management, and autonomous systems design) and the ITU (which has historically played a role in standards for information and communications technologies) are of relevance. Some of these bodies are not intergovernmental, but private, and are respected and abided.

International standards bodies can serve as focal points through which opposing perspectives can be reconciled, providing a common governance framework from which to build further agreement. Moreover, setting successful standards will require

2. The US Executive Order on “Maintaining American Leadership in Artificial Intelligence” identified international standards as a priority (NIST 2019). Likewise, in 2018, China Electronics Standardization Institute, within the Ministry of Industry and Information Technology, launched an “AI Standardization White Paper” advising the government to promote a set of universal regulatory principles and standards to ensure the safety of AI technology.

3. Until now, the field has mostly been dominated by the United States, Europe, and Japan, but China’s ability to transform this landscape is also expanding with the growth in its capacity to propose core innovations in a growing number of emerging technological fields, such as 5G, but clearly involving AI.

coordinated efforts from the AI community (within the private sector, academia, governments, and international and transnational bodies), external stakeholders, and engagement groups (such as the Think 20 and Business 20). This would enable the active promotion of their development and use, since public, private, and research institutions need to be included in the discussion.

International treaties, national requirements, and other global pressures can also contribute to the dissemination of standards once they have been established. For instance, national regulations can allude to existing standards and instruct compliance *de jure*. Meanwhile, the World Trade Organization's Agreement on Technical Barriers to Trade mandates WTO members to use international standards where they exist and are effective and appropriate. Likewise, the leadership of the G20 could play a key role in providing an initial roadmap toward global solutions where national rules may fall short. To this end, the G20 could discuss potential standards and their posterior commitment to promoting their adherence globally.

Proposal IV

Ensure accountability and transparency and foster legitimacy

Building common understanding, designing standards, and empowering oversight authorities are fundamental steps to curtail the challenges that AI systems entail. However, many have stated that when public authorities delegate decision-making powers to AI, a series of gaps emerge. This is especially pronounced in the case of automated decision systems. These are data-driven tools used to analyze datasets and generate scores, predictions, classifications, or some recommended actions deployed to make decisions that impact human welfare (Richardson 2019). It is the integrity of the procedures as a whole—their transparency, accountability, and legitimacy—which would ensure that AI works for all and that people trust public policy.

First, standards should not only be discussed and implemented by technocrats and specialists. Given the potential and observed impact of these systems on the lives and rights of people, there is a need for a commitment from G20 leaders to make information publicly available. This would allow the public to meaningfully assess how systems implemented for public policy function and how they are being used, thereby ensuring transparency and empowering people.

Second, organizations that develop software must be able to demonstrate due diligence in their creation, documenting the codes that are written (stating by whom, when and why) and which software and data libraries are being used. They must also

undertake appropriate testing before the software are released and monitor them while the code is in use (Bryson 2018). This would allow for transparency as well as accountability.

Finally, the G20 countries should adopt clear procedures relating to the collection, usage, storage, and sharing of personal information in the context of developing and implementing a given AI system in a privacy-preserving manner, allowing for informed consent.

To achieve these policy options, online global public consultations and other innovative methods could be put in place to help participation and information dissemination.

The role of the G20

AI is a growing international challenge that requires coordinated global responses. This does not imply, however, that all technology governance must be global. Regions, states, and cities must be able to respond to the specific social, economic, and cultural demands of their citizens. However, the most important principle is interoperability. Therefore, defining comparable global standards will be decisive in managing the digital transition. Many of the corporations creating AI systems operate across national boundaries; as a result, geopolitical and cultural differences would arise. Therefore, we will require institutional guidelines for transnational coordination. As such, standards can provide the infrastructure needed to develop new technologies, as well as the required safety procedures to do so in a controlled manner, making sure AI and automated decision systems work for many.

The role of the G20 in aligning interests and leading such processes will be crucial in stimulating establishment, adoption, and dissemination of standards, as it helps them grow in influence. As a central forum for debate and dialogue that brings together the main political and economic forces of the world, the G20 is the best platform to lead the conversation on the future of digital governance and respond to one of the biggest challenges our world is facing today. The involvement of the international community and experts in this process will contribute to the development of better, transparent, and legitimate standards. By engaging in this debate, the G20 has the potential to become a leading space of dialogue for creating a new architecture for the 21st century. It would thus manage the transition to a multi-stakeholder AI governance, and ensure a better future for all.

Disclaimer

This policy brief was developed and written by the authors and has undergone a peer review process. The views and opinions expressed in this policy brief are those of the authors and do not necessarily reflect the official policy or position of the authors' organizations or the T20 Secretariat.



REFERENCES

Austin, John. 1995. Frontmatter. In *The Province of Jurisprudence Determined*, edited by Wilfred E. Rumble (Cambridge Texts in the History of Political Thought, pp. I–IV). Cambridge: Cambridge University Press.

Baldwin, Robert, Martin Cave, and Martin Lodge. 2012. *Understanding Regulation: Theory, Strategy, and Practice*. Oxford University Press on Demand. <https://doi.org/10.1093/acprof:osobl/9780199576081.001.0001>.

Bostrom, Nick. 2014. *Superintelligence*. Oxford: Oxford University Press.

Bryson, Joanna. 2018. “AI & Global Governance: No One Should Trust AI.” *AI & Global Governance Articles & Insights*. Accessed June 8, 2020. <https://cpr.unu.edu/ai-global-governance-no-one-should-trust-ai.html>.

Cihon, Peter. 2018. “Regulatory Dynamics of Artificial Intelligence Global Governance.” Typhoon Consulting.

Crawford, Kate, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Amba Kak, et al. *AI Now 2019 Report*. New York: AI Now Institute, 2019. https://ainowinstitute.org/AI_Now_2019_Report.html.

European Commission. 2019. “Ethics Guidelines for Trustworthy Artificial Intelligence.” High-Level Expert Group on AI. Last modified April 8, 2019. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

Fjeld, Jessica, and Adam Nagy. 2020. “Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI.” Accessed April 6, 2020. <https://cyber.harvard.edu/publication/2020/principled-ai>.

IEEE (Institute of Electrical and Electronic Engineers). 2019. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design*. Available at: <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf>

Imbrie, Andrew, and Elsa B. Kania. 2019. *AI Safety, Security, and Stability Among Great Powers Options, Challenges, and Lessons Learned for Pragmatic Engagement*. CSET Policy Brief. Accessed April 7, 2020. <https://cset.georgetown.edu/wp-content/uploads/AI-Safety-Security-and-Stability-Among-the-Great-Powers.pdf>.

REFERENCES

Mittelstadt, Brent. 2019. "Principles Alone Cannot Guarantee Ethical AI." *Nature Machine Intelligence*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3391293.

Mulgan, Geoff. 2018. *Big Mind: How Collective Intelligence Can Change Our World*. Nueva Jersey, Princeton University Press.

NIST (National Institute of Standards and Technology). 2019. *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools Prepared in response to Executive Order 13859*. Submitted on August 9, 2019. https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

Pomares, Julia, and María B. Abdala. 2020. "The Future of AI Governance: The G20's Role and the Challenge of Moving Beyond Principles." *Global Solutions Journal Issue 5*. https://www.global-solutions-initiative.org/wp-content/uploads/2020/04/GSJ5_Pomares_Abdala.pdf.

Richardson, Rashida, ed. 2019. *Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*. AI Now Institute. Accessed April 7, 2020. <https://ainowinstitute.org/ads-shadowreport-2019.html>.

Russell, Stuart, and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach*. Third Edition, Pearson.

Ortega, Andrés. 2020. "Geopolítica de la ética en Inteligencia Artificial." Documento de trabajo. Real Instituto Elcano. Accessed April 7, 2020. http://www.realinstitutoelcano.org/wps/portal/rielcano_es/contenido?WCM_GLOBAL_CONTEXT=/elcano/elcano_es/zonas_es/dt1-2020-ortega-geopolitica-de-la-etica-en-inteligencia-artificial.

Seaman, John. 2020. "China and the New Geopolitics of Technical Standardization." Notes de l'Ifri. Accessed April 4, 2020. <https://www.ifri.org/en/publications/notes-de-lifri/china-and-new-geopolitics-technica/lstandardization>

Selznick, Philip. 1985. "Focusing Organizational Research on Regulation." In *Regulatory Policy and the Social Sciences*, edited by Roger Noll, 363–364. Berkeley: University of California Press.



AUTHORS

María Belén Abdala

CIPPEC Argentina

Andrés Ortega

Elcano Royal Institute

Julia Pomares

CIPPEC Argentina

