Learning from the past for designing AI global governance systems

by Jacob Greenspon

A pre-read paper from the Schwartz Reisman Institute for Technology and Society

Innovating AI Governance: Shaping the Agenda for a Responsible Future
December 2020

Table of Contents

1. Abstract

Rapid developments in AI technologies have brought to the forefront questions of whether, and how, we can regulate AI. Given the globalized nature of AI, some form of international regulatory arrangement is necessary. This echoes the motivation behind several historic regulatory challenges, including notably the post-war Bretton Woods system of international economic cooperation as well as the 1975 Asilomar Conference on recombinant DNA and, more recently, the Maritime Labour Convention. These and other past exemplars are analyzed to understand how these regulatory institutions were shaped by the attributes of the regulatory challenge and the choices among tradeoffs of global governance. This sheds light on how these challenges can be navigated when contemplating global regulatory cooperation on AI.

2. Background: Recent developments in AI governance and regulation

Artificial intelligence has developed by leaps and bounds over the past decade. AI technologies has been used to improve cancer diagnosis through CT scans, prevent illegal fishing, and boost profits of a diverse set of firms.[1] These developments, and more on the horizon, offer optimism for the technological, social, and other challenges that AI can help solve.

However, the rapid development of AI has also shed light on the problems they may bring. One is job displacement: some fear that recent machine learning-based GPT-3 technology that can draft emails, design websites, and write code will eliminate the need even for executive assistants, web developers, and programmers. Another concern is the increasing documentation of 'algorithmic bias', such as facial recognition systems that disproportionately create false matches for people of colour or the COMPAS algorithms used in bail hearings and sentencing decisions across the US.[2] Even when performing exactly as designed, AI technologies can lead to unintended—or even illegal—outcomes, for example collusion between consumer pricing algorithms.[3] Future technological developments will likely lead to the discovery of further undesirable consequences of AI.

In response to these and other concerns, several initiatives have been created to govern and regulate AI. National governments have set up advisory bodies to study the ethical use of AI, such as the Singapore Advisory Council on the Ethical Use of AI and Data and the Austrian Council on Robotics and AI. Some have taken action to regulate specific AI technologies, for

[1] "Sustainability through Transparency," Global Fishing Watch, accessed November 9, 2020, https://globalfishingwatch.org/; "How AI And Deep Learning Are Now Used To Diagnose Cancer," Infervision, accessed November 9, 2020, https://global.infervision.com/news/48.html; Jacques Bughin et al., "Artificial Intelligence - The Next Digital Frontier?," Discussion Paper (McKinsey Global Institute, June 2017), https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx.
[2] James Zou and Londa Schiebinger, "AI Can Be Sexist and Racist — It's Time to Make It Fair," Nature 559, no. 7714 (July 2018): 324–26, https://doi.org/10.1038/d41586-018-05707-8; Jack Clark and Gillian K. Hadfield, "Regulatory Markets for AI Safety," ArXiv:2001.00078 [Cs, Econ, q-Fin], December 11, 2019, http://arxiv.org/abs/2001.00078.
[3] Emilio Calvano et al., "Artificial Intelligence, Algorithmic Pricing, and Collusion," American Economic Review 110, no. 10 (October 2020): 3267–97, https://doi.org/10.1257/aer.20190623.

example through the 2019 Experimental Law on Selfdriving vehicles in the Netherlands, or the Canadian Directive on Automated Decision-Making.[4] Intergovernmental organizations have also adopted principles related to the responsible development of AI; in 2019 alone, the OECD, G20, European Union and World Economic Forum each adopted a set of AI principles (with varying degrees of scope).[5] Several commercial entities have also adopted principles, such as a set of Responsible AI Practices released by Google and the OpenAI Charter.[6] Finally, non-governmental standard-setting membership organizations such as the ISO and IEEE have launched AI initiatives that focus on developing standards.[7]

The existing and proposed initiatives to govern the development of AI span a wide range of possible mechanisms. Some are based on governments setting and enforcing rules, while others are based instead on self-enforced guidelines determined by commercial entities (who have a clear conflict of interest). Other models such as 'regulatory markets' leverage the power of market-based incentives to encourage licensed companies to develop third-party regulatory technologies that will ensure regulated entities meet outcomes determined by governments.[8] Some form of regulation is necessary to ensure the safe and beneficial development of AI. This is in part due to the competitive race dynamics of development: private firms want to develop better AI faster than their rivals, and it is costly to ensure these developments are safe.[9] Each approach has benefits and drawbacks for dealing with this collective action problem, and some are better fit to certain circumstances than others. Nonetheless, each must contend with the global nature of AI.

3. The globalized context of AI: R&D, MNCs, and national security

An essential characteristic of AI technologies is that they are developed and adopted across borders. AI has developed during a period of immense globalization, and its intangible nature

[4] Ministerie van Infrastructuur en Waterstaat, "Green light for Experimental Law for testing self-driving vehicles on public roads - News item - Government.nl," nieuwsbericht (Ministerie van Algemene Zaken, July 2, 2019), https://www.government.nl/latest/news/2019/07/02/green-light-for-experimental-law-for-testing-self-driving-vehicles-on-public-roads; Treasury Board of Canada Secretariat, "Directive on Automated Decision-Making," February 5, 2019, https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

[5] "OECD Principles on Artificial Intelligence - Organisation for Economic Co-Operation and Development," accessed November 9, 2020, https://www.oecd.org/going-digital/ai/principles/; "G20 Ministerial Statement on Trade and Digital Economy" (Group of 20, June 9, 2020), https://www.mofa.go.jp/mofaj/files/000486596.pdf; "On Artificial Intelligence - A European Approach to Excellence and Trust" (European Commission, February 19, 2020), https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf; "World Economic Forum Guidelines for AI Procurement" (World Economic Forum, September 2019), http://www3.weforum.org/docs/WEF_Guidelines_for_AI_Procurement.pdf.

[6] "Google Responsible Development of AI" (Google, n.d.), https://ai.google/static/documents/responsible-development-of-ai.pdf; "OpenAI Charter," OpenAI, accessed November 9, 2020, https://openai.com/charter/.

[7] "The New Frontier for Artificial Intelligence," ISO, accessed November 9, 2020, https://www.iso.org/cms/render/live/en/sites/isoorg/contents/news/2018/10/Ref2336.html; "IEEE Global Initiative for Ethical Considerations in Artificial Intelligence (AI) and Autonomous Systems (AS) Drives, Together with IEEE Societies, New Standards Projects; Releases New Report on Prioritizing Human Well-Being," accessed November 9, 2020, https://standards.ieee.org/news/2017/ieee_p7004.html.

[8] Clark and Hadfield, "Regulatory Markets for AI Safety."

[9] Clark and Hadfield.

reduces the importance of borders (especially with the growth of cloud-based storage and processing). There are three aspects of this globalized context that are particularly relevant to the regulation and governance of AI.

a. AI research and development activities are globally distributed and linked

First, research and development (R&D) related to AI increasingly occurs across the world.[10] For example, between 2012 and 2017 there was substantial internationalization of participation at the Association for the Advancement of Artificial Intelligence conference, with high growth among authors affiliated with East Asian countries in particular.[11] The globalized nature of AI R&D is in some ways unsurprising. While researchers in the US and UK were responsible for early developments, the relatively low barriers to entry for developing AI technologies (compared to, for example, advanced biochemicals or medical instruments) makes it easy to replicate and build on these elsewhere.[12] Cloud computing "helps democratize innovation" by minimizing the time, money, and resources required for developing AI.[13] AI-related graduate programs often attract students from around the world, many of whom return home afterwards. Researchers often collaborate and travel internationally for specific projects. For example, US-based researchers are collaborating in 12% of deep learning projects under the auspices of the EU Horizon 2020 research framework, as well as 4% of machine-learning related projects.[14] This occurs in the private sector as well: for instance, Uber, Samsung, NVIDIA, LG Electronics, and many other multinational enterprises have established AI-specific R&D centres across Canada.[15]

b. Multinational corporations play a major role in AI

This last point highlights another important globalized aspect of AI: that the firms developing AI have staff, investors, and customers located across the world. Multinational corporations often locate AI research units in various 'hub' regions and cities (Silicon Valley, Berlin, Seattle, London, Shanghai, Boston, Toronto, and Montreal) around the world "where the talent is, rather than forcing the talent to move to where the company is."[16] These firms rely on investment

[10] Raymond Perrault et al., "Artificial Intelligence Index: 2019 Annual Report" (Stanford, CA: AI Index Steering Committee, Human-Centered AI Institute, Stanford University, December 2019), https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf.
[11] Avi Goldfarb and Daniel Trefler, "AI and International Trade" (National Bureau of Economic Research, January 29, 2018), https://doi.org/10.3386/w24254.
[12] "The History of Artificial Intelligence," Science in the News (blog), August 28, 2017, http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.
[13] Alexander Benlian et al., "Special Section: The Transformative Value of Cloud Computing: A Decoupling, Platformization, and Recombination Theoretical Framework," Journal of Management Information Systems 35, no. 3 (July 3, 2018): 719–39, https://doi.org/10.1080/07421222.2018.1481634.
[14] Christie Lawrence and Sean Cordey, "The Case for Increased Transatlantic Cooperation on Artificial Intelligence," The Cyber Project (Belfer Center for Science and International Affairs, Harvard Kennedy School, August 2020), https://www.belfercenter.org/sites/default/files/2020-08/TransatlanticAI.pdf.
[15] "Canada's AI Ecosystem: Government Investment Propels Private Sector Growth" (University of Toronto, Government Relations Office, n.d.), https://d3n8a8pro7vhmx.cloudfront.net/uot/pages/301/attachments/original/1594219597/GRO_AI_Report_FINAL_2.pdf?1594219597.
[16] Goldfarb and Trefler, "AI and International Trade."

capital from around the world, for example from Japan's SoftBank Vision Fund 2, which includes investors from the US, Taiwan, Kazakhstan, and elsewhere.[17] Finally, most companies offering AI-based products rely on international sales. While the total value of AI exports is difficult to measure directly, the global customer base of several large companies reliant on AI technologies—such as Facebook, Amazon, Alibaba, and others—suggests that cross-border flows of AI-based products are not negligible. AI is also used to improve trade itself by making global value chains more efficient, such as through better predicting future trends in consumer demand and supply stocks and improving risk management.[18]

c. National security considerations are increasingly important for AI

Finally, the development of AI is increasingly discussed in terms of national security. Most prominent are the frequent analogies to an AI 'arms race' between the US and China.[19] There has been substantial state investment in AI-powered military technologies for intelligence collection and analysis, cyber operations, autonomous vehicles, and more.[20] Even the armed forces of less developed countries and non-state actors have used AI technologies, for example to conduct "swarming attacks" of automated drones.[21] But, as recent discussions of the production of advanced computer chips suggests, the links between AI and national security extend beyond military applications.[22] Many AI innovations will have potential applications for surveillance or the military, and as these innovations are increasingly viewed as a source of strength—or threat—for national security, their development will be characterized as adversarial rather than collaborative.

One aspect of AI that illustrates this globalized context is data. Existing trade agreements such as the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) and the United States-Mexico-Canada Agreement (USMCA) include commitments for the free flow of data.[23] These trade protections are designed in part to allow multinational firms to leverage user data from around the world in the development of AI-based products. This user data is collected on servers outside a nation's borders, which creates potential threats to both privacy and national security if the data is not deemed to be adequately protected. However, these international data flows create challenges for domestic privacy laws that supposedly stop at a country's borders.

---

[17] Sam Shead, "SoftBank Launches New $108 Billion Vision Fund To Invest in AI," Forbes, accessed November 9, 2020, https://www.forbes.com/sites/samshead/2019/07/26/softbank-launches-new-108-billion-vision-fund-to-invest-in-ai/.
[18] Joshua P. Meltzer, "The Impact of Artificial Intelligence on International Trade," Brookings (blog), December 13, 2018, https://www.brookings.edu/research/the-impact-of-artificial-intelligence-on-international-trade/.
[19] Ryan Hass and Zach Balin, "US-China Relations in the Age of Artificial Intelligence," Brookings (blog), January 10, 2019, https://www.brookings.edu/research/us-china-relations-in-the-age-of-artificial-intelligence/.
[20] Kelley M Sayler, "Artificial Intelligence and National Security" (Congressional Research Service, August 26, 2020), https://fas.org/sgp/crs/natsec/R45178.pdf.
[21] James Johnson, "Artificial Intelligence, Drone Swarming and Escalation Risks in Future Warfare," The RUSI Journal 165, no. 2 (February 23, 2020): 26–36, https://doi.org/10.1080/03071847.2020.1752026.
[22] Saif M. Khan and Carrick Flynn, "Maintaining China's Dependence on Democracies for Advanced Computer Chips," Brookings (blog), April 27, 2020, https://www.brookings.edu/research/maintaining-chinas-dependence-on-democracies-for-advanced-computer-chips/.
[23] Meltzer, "The Impact of Artificial Intelligence on International Trade."

In general, while there are many benefits to AI being so globalized—such as the more rapid development and diffusion of useful technologies—it creates complications for its regulation and governance. Fortunately, AI is not entirely unique in terms of the globalized nature of issues related to its regulation. Several other technologies have emerged in a similar context, and many other issues are defined by their international nature. Past global regulatory cooperation in these domains sheds light on how approaches to regulating AI can navigate these global challenges.

4. Exemplars of global governance and regulation

We can learn from the past to chart a future path of global AI governance and regulation. There have been several cases of global arrangements that sought to address some regulatory challenge in the most socially beneficial way. While most of these represent successes of global governance, some (such as on cryptography) are more limited. The following examples will be further developed in later sections that discuss key issues and tradeoffs in the design of global governance and regulatory institutions.

a. Asilomar Conference (International Conference on Recombinant DNA Molecules), 1975

This was an international meeting of biologists, lawyers, and physicians (in Asilomar, California) to discuss regulation of biotechnology, particularly recombinant DNA (combining DNA from different organisms). It was preceded by a voluntary moratorium on recombinant DNA research in academic and industrial research centres, in response to biohazard concerns. The meetings created a system of "assign[ing] a risk estimate to the different types of experiments envisaged, and apply[ing] safety guidelines of varying stringency according to the degree of risk."[24] It also prohibited research experiments that involved biohazards that were not technologically possible to contain at the time. These formed the basis of the official US guidelines on recombinant DNA issued a year later.[25] The conference allowed this science to develop in a safe and socially beneficial way and is also credited with generally restoring public trust in science by bringing science policy discussions into the open.

b. Bretton Woods system and the International Monetary Fund (IMF)

The Bretton Woods system was the establishment of rules and institutions—in particular, the International Monetary Fund (IMF)—to promote international monetary integration. In his closing address at the 1944 Bretton Woods Conference, US Treasury secretary Henry Morgenthau Jr spoke of having "come to recognize that the wisest and most effective way to protect our national interests is through international co-operation — that is to say, through united effort for the attainment of common goals."[26] While the IMF was the centerpiece of this

[24] Paul Berg, "Asilomar 1975: DNA Modification Secured," Nature 455, no. 7211 (September 2008): 290–91, https://doi.org/10.1038/455290a.
[25] Berg.
[26] "Opinion: Time to Get Global Co-Operation on the Agenda," accessed November 9, 2020, https://www.theglobeandmail.com/business/commentary/article-time-to-get-global-co-operation-on-the-agenda/.

system, the World Trade Organization (see below) and World Bank were two other important pillars of Bretton Woods.

The most important mechanism of this monetary cooperation was a "par value system" of exchange rates where nations kept the value of their currencies pegged at agreed-upon rates (that were also linked to the price of gold).[27] These exchange rates could only be adjusted to correct a "fundamental disequilibrium" in the balance of payments between nations, and only with the agreement of the IMF (which was governed by the member states). While there had previously been a gold standard for the decades before and after World War I, the Bretton Woods system was the first fully negotiated monetary order between independent nations. The goal was to avoid repeat of the economic policies such as increased trade protectionism, competitive currency devaluations, and limits on foreign-currency holdings that had contributed to the Great Depression. This system flourished for nearly three decades, coinciding with post-war economic expansions across member nations and generally peaceful international economic relations. However, the Bretton Woods system fell apart starting in the late 1960s as the US dollar was seen as increasingly overvalued, and then the gold standard dissolved in August 1971 when the US temporarily suspended the convertibility of US dollars to gold. Following a short-lived attempt to revive fixed exchange rates, since 1973 most major currencies have instead 'floated' against each other.

c. European Union Single Market Initiative (EU SMI)

The EU Single Market is a set of rules, principles, and treaties that seek to ensure that people, goods, services, and money can move freely throughout the participating economies. It is intended to increase competition and specialization, leverage economies of scale, and improve the efficiency of resource allocation across the EU (and participating non-EU states such as Norway, Switzerland, Iceland, and the UK). This is achieved through harmonizing standards (especially for trade in goods), prohibiting border levies, replacing customs controls by audits and risk analysis, ensuring fairness in taxation, and eliminating restrictions on capital flows. It also allows for the free movement of workers, including their eligibility for some social benefits. In some cases, the Single Market requires harmonization of national standards, while in others it operates under the mutual recognition principle whereby roughly equivalent standards in each country are recognized as sufficiently harmonized without needing further modification.

d. Forest Stewardship Council (FSC)

The FSC is an international non-profit organization established in 1993 to promote responsible forest management through market-based certification. It was created in response to widespread deforestation and the need for alternative governance arrangements in the face of the failure of national governments to reach consensus on regulation. It is composed of forest owners, timber firms, and NGOs that collaborate to develop global as well as national standards

[27] "About the IMF: History: Cooperation and Reconstruction (1944–71)," accessed November 9, 2020, https://www.imf.org/external/about/histcoop.htm.

(adapted to local conditions) for responsible forest management. The FSC then accredits certification bodies that evaluate and monitor companies for compliance with these standards.

e. International Organization for Standardization (ISO)

The International Organization for Standardization (ISO) is a non-governmental organization responsible for developing common, but voluntary technical standards in order to facilitate world trade and integration for a wide range of products and services. These standards promote product compatibility and safety.[28] Since 1946, nearly 800 technical committees from 165 member nations have developed over 23,000 international standards "covering almost all aspects of technology and manufacturing."[29] For example, the ISO has developed standards for day-to-day items such as paper sizes and formats, as well as for more sensitive socio-technical issues like cybersecurity and nuclear energy. These standards are crafted on a consensus-basis by committees of technical experts (from academia, industry, government, and elsewhere) who are nominated by national member organizations, which also approve standards by voting.[30] The standards are then enforced by other existing institutions through de facto methods like governments exercising their purchasing power as well as through de jure methods such as standardization mandates under World Trade Organization agreements.[31]

The ISO has so far developed at least 46 standards relevant to responsible AI. This includes standards on the assessment of machine learning classification performance, addressing biases in AI systems, and big data standards.[32] However, some argue that while standards are helpful for influencing the development and deployment of AI through product specifications, building trust among key actors, and disseminating best practices globally, "existing market forces are insufficient to incentivize the adoption of standards that govern fundamental research and other transaction-distant systems and practices."[33]

f. Maritime Labour Convention, 2006 (MLC)

The Maritime Labour Convention is a set of international treaties under the auspices of the International Labour Organization, a UN agency. It combines and updates many existing maritime labour standards. Since taking effect in 2012, the MLC has established minimum working conditions for seafarers, such as regarding their hours of work and rest,

---

[28] "ISO in Brief" (International Organization for Standardization, 2018), https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100007.pdf.
[29] "ISO - About Us," ISO, accessed November 9, 2020, https://www.iso.org/about-us.html.
[30] "ISO in Brief."
[31] Peter Cihon, "Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development" (Center for the Governance of AI, Future of Humanity Institute, University of Oxford, April 2019), https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf.
[32] Review of AI-related standards from AI Global. https://docs.google.com/spreadsheets/d/12R4ztw7Ewz5KIGMWYFIy1epZWtP-s0bMMdIx0FGftpg/edit#gid=0
[33] Cihon, "Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development."

accommodation, health protection, medical care, welfare and social security protection.[34] As of 2019 the MLC has been signed by 97 countries that represent 91% of global shipping. Even ships from non-signatory countries must meet these requirements, however, since ratifying states must enforce sanctions on any ships coming into their harbors according to a "no more favorable treatment principle."[35] The MLC was drafted and is under continuous review by a tripartite process involving governments, ship owners, and seafarers unions.

g. Medical Device Single Audit Program (MDSAP)

The MDSAP, which became operational in 2017, was created to facilitate the development and adoption of medical devices (everything from tongue depressors to pacemakers) across national borders. Specifically, its goal was to "to create a scheme in which a medical device can be audited by a single organization for compliance with the (quality management) standards of any of the countries in which it will be sold. Countries participating in the program agree to accept the audit report of the single auditor as meeting the certification requirements of their regulatory scheme."[36] This is a version of a regulatory market: private sector auditors compete to provide auditing services to device manufacturers, and are regulated by governments. Importantly, it creates a global market for regulation in which individual countries are not obliged to adopt the same regulatory standards. Instead, manufacturers can pass a single audit and be deemed compliant with regulations across all the participating national regulators (Australia, Brazil, Canada, Japan and the U.S.).

h. OECD Guidelines for Cryptography Policy, 1997

Prior to the adoption of the OECD Guidelines for Cryptography Policy in 1997, there was very limited international cooperation around cryptography (defined generally as the code-based transformation of information so that it cannot be understood without some decryption key).[37] Nations tended to pursue cryptography as a tool for national security and attempted to use various forms of domestic regulation and export controls to limit the diffusion of cryptographic technology.[38] The 1997 OECD Guidelines, which were subsequently adopted by several countries through national regulation, marked a shift towards an international "relaxation of regulations concerning encryption."[39] However, the absence of global regulation has contributed

---

[34] "Milestone Ratifications of Seafarers' Labour Rights Charter," News, August 20, 2012, http://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_187660/lang--en/index.htm.
[35] "Maritime Labour Convention, 2006 (MLC, 2006), Frequently Asked Questions (FAQ), Online Revised Edition, 2012" (International Labor Organization, 2012), https://www.ilo.org/wcmsp5/groups/public/---ed_norm/---normes/documents/publication/wcms_177371.pdf.
[36] Clark and Hadfield, "Regulatory Markets for AI Safety."
[37] Stewart A. Baker, "Decoding OECD Guidelines for Cryptography Policy," The International Lawyer 31, no. 3 (1997): 729–56.
[38] Jade Leung, "Who Will Govern Artificial Intelligence? Learning from the History of Strategic Politics in Emerging Technologies" (DPhil Thesis, University of Oxford, 2019), https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665.
[39] "OECD Guidelines for Cryptography Policy - OECD," accessed November 9, 2020, https://www.oecd.org/sti/ieconomy/guidelinesforcryptographypolicy.htm; W Schulz and J van Hoboken, Human Rights and Encryption (United Nations Educational, Scientific, and Cultural Organization, 2016); Wayne Madsen and David Banisar,

to a fragmentation of the global market into several non-interoperable cryptographic technologies.[40] Moreover, there has been little action on essentially international issues related to cryptography and human rights.

i. Outer Space Treaty, 1967

The Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space was signed by the US, UK, and USSR in 1967 to ensure the peaceful exploration of space.[41] It involved a rare agreement among adversaries to limit the use of an evolving technology, holding that nuclear weapons would not be allowed in space and that celestial bodies could not be colonized or used for military purposes. The treaty is deliberately short and was designed to be flexibly interpreted so that issues that arose later in space exploration would be addressed. This also allowed a deal to get done in time: "Not letting the best be the enemy of the good meant that by the time man landed on the moon we had a global political framework as a foundation on which to build."[42] It continues to be in force and has now been signed by 104 nations.

j. United Nations (UN)

The United Nations was founded in 1945, at the conclusion of World War Two, to maintain international peace and security and achieve international cooperation. There are three main UN bodies relevant to this discussion of global AI governance. The UN General Assembly (UNGA), which is the main deliberative body of the UN, meets in annual sessions of all UN member states and approves most resolutions (which are non-binding) by a simple majority vote. The UNGA can make recommendations on any matters within the scope of the UN with the exception of issues related to peace and security. The UN Security Council (UNSC) is able to make binding decisions (known as resolutions) on member states in order to maintain international peace and security. The UNSC is composed of 15 member states—five that are permanent members and can veto a resolution, and ten that are elected to two-year terms voted by the UNGA. Finally, the UN Economic and Social Council (ECOSOC) promotes international cooperation, largely by coordinating the work of a diverse set of UN agencies. ECOSOC has 54 members elected by the UNGA for three-year terms.

"Cryptography and Liberty 2000: An International Survey of Encryption Policy" (Washington, D.C.: Electronic Privacy Information Center, 2000), https://epic.org/reports/crypto2000.html/.

[40] Leung, "Who Will Govern Artificial Intelligence? Learning from the History of Strategic Politics in Emerging Technologies."

[41] "Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies," United Nations Office for Disarmament Affairs, accessed November 9, 2020, http://disarmament.un.org/treaties/t/outer_space/text.

[42] Verity Harding, "Lessons from History: What Can Past Technological Breakthroughs Teach the AI Community Today," accessed November 9, 2020, https://www.bennettinstitute.cam.ac.uk/blog/lessons-history-what-can-past-technological-breakt/.

k. World Trade Organization (WTO)

The WTO was officially created in 1995 to set and enforce rules for international trade, provide a forum for negotiating and monitoring trade liberalization, and increase the transparency of decision-making processes around trade. This importantly involves governing countries' use of trade protection measures (such as tariffs and standards) and, in particular, disputes around these measures. For the WTO, dispute resolution relies on a multi-stage process of consultations, expert panel reports, hearings, and appeals to an adjudicative Dispute Settlement Appellate Body, which can sanction retaliatory trade actions.[43] However, recently opposition from the US has limited the operations of this Appellate Body.[44]

The WTO was a successor to the General Agreement on Tariffs and Trade (GATT), which had been in force since 1948. The GATT was designed to boost economic growth by facilitating a reduction in trade barriers, limiting discriminatory tariff preferences, and providing a system for peacefully resolving trade disputes.[45] It was created as one of three pillars (along with the International Monetary Fund and the World Bank) of the Bretton Woods system, whereby the victorious Allied nations designed the governance of international monetary and commercial relations after World War Two. As with the WTO, GATT relied on reciprocity in trade protections, multilateral negotiations, and nondiscrimination in trade (the extension of trade concessions to all members).

[43] "Understanding the WTO - Settling Disputes," accessed November 9, 2020, https://www.wto.org/english/thewto_e/whatis_e/tif_e/disp1_e.htm.
[44] Keith Johnson, "How Trump May Finally Kill the WTO," Foreign Policy (blog), accessed November 9, 2020, https://foreignpolicy.com/2019/12/09/trump-may-kill-wto-finally-appellate-body-world-trade-organization/.
[45] Douglas A. Irwin, "The GATT in Historical Perspective," The American Economic Review 85, no. 2 (1995): 323–28.

5. Summary table of exemplars and framework of global regulatory arrangements

The table below explores how each of the exemplars fits into a framework of how the attributes of a regulatory challenge influence the design of governance systems given several tradeoffs (discussed below). It also summarizes how exemplar regulatory challenges compare to the AI regulatory challenge. The entries are suggestive, primarily intended to provide a structure for discussion of the global regulatory challenges around AI.

Exemplar	Non-state involvement	Barriers to entry	National security concerns	Ability to avoid or opt-out of regulation	Global integration trilemma	Compliance vs. comprehensiveness of membership	Accountability vs. ability to agree	Comparability to AI
Asilomar Conference	Mostly public, some private and civil society	High	Moderate	High	Limits national sovereignty	Compliance to regulations	Ability to agree	Similar cases of rapidly developing technologies causing public concern, but different actors involved
EU Single Market	Public sector (govts)	Low	Low	Low	Limits national sovereignty	Compliance to regulations	Ability to agree	Similar focus on standards harmonization, but differ on actors involved and national security considerations
Bretton Woods and IMF	Public sector (govts)	High	Moderate	Moderate	Limits international integration	Compliance to regulations	Accountability	Similar impetus to come together to create global regulatory system, but different actors involved
Forest Stewardship Council	Private and civil (govts as land-owners but not members)	Moderate	Low	High	Limits public purpose	Comprehensiveness of membership	Ability to agree	Similar in role for non-state actors and global implications with interest in harmonization, differ on speed of technology and complexity, national security considerations
International org. for Standardization	Private, public, and other groups very involved	Varies based on product	Varies based on product; typically low	High (voluntary standards)	Limits public purpose	Comprehensiveness of membership	Ability to agree	Similar variety of actors involved and ability to opt-out, differs on national security considerations

Exemplar	Non-state involvement	Barriers to entry	National security concerns	Ability to avoid or opt-out of regulation	Global integration trilemma	Compliance vs. comprehensiveness of membership	Accountability vs. ability to agree	Comparability to AI
Maritime Labor Convention	Public, private, and civil society (tripartite process)	Moderate	Low	Moderate	Limits national sovereignty	Comprehensiveness of membership	Ability to agree	Similar in role for non-state actors and global implications with interest in harmonization, differ on speed of technology and complexity, national security considerations
MDSAP	Public and private sectors	High	Low	Moderate	Limits public purpose	Compliance to regulations	Ability to agree	Similar in speed of technology and global reach, role of non-state, differ on scope of innovation
OECD Cryptography guidelines	Public, private, and civil society	Moderate	High	High	Limits international integration	Comprehensiveness of membership	Ability to agree	Similar variety of relevant actors, national security considerations, and ability to opt-out; differ on barriers to entry
Outer Space Treaty	Public sector (govs.)	Very high	High	Moderate	Limits national sovereignty	Comprehensive	Accountability	Similar on national security and global implications, differ on role of non-state actors and speed/scope of innovation
United Nations	Mostly public, some private and civil society	N/A	High	Low	Limits national sovereignty	Comprehensive (UNGA), Compliance (UNSC)	Accountability	Similar on global reach but differ on role of non-state and speed and complexity of technology
WTO/ GATT	Mostly public, some private and civil society	Moderate	Moderate	Low	Limits national sovereignty	Comprehensive (GATT), Compliance (WTO)	Accountability	Similar on broad global scope and value of harmonization but differ on speed of technological change

6. Key attributes of a regulatory challenge inform the design of its global governance system

Each of the above institutions was designed to address a certain regulatory challenge with global ramifications. These governance and regulatory arrangements were therefore shaped in part by the inherent characteristics of the regulatory domain. There are four dimensions on which each regulatory challenge varies that are particularly influential to the design of the governance arrangement: involvement of non-state actors, barriers to entry, national security considerations, and availability of outside options. When analyzed along these four dimensions, the above exemplars illustrate some general lessons for the design of global governance and regulatory systems. This sheds light on several issues likely to be encountered when designing global AI systems specifically and provides insights into how these challenges may be navigated.

a. What is the degree of involvement by non-state actors?

The institutions designed to address emergent global issues typically involve engagement by actors other than nation states: the private sector and civil society (including non-governmental organizations). The domain that is being regulated and the period during which a governance system is created explain a large degree of the variance in the extent to which private actors are involved in the design and enforcement of global regulation. The Outer Space Treaty, for example, regulated a domain where states had at that time a monopoly on the technology in question. Private firms and civil society organizations played little role in the Treaty's creation, despite the fact that it held countries responsible for non-governmental activities in space, since it was initially signed decades before commercial activities in space were seriously contemplated.[46] The private sector had little involvement with the United Nations—whose very name suggests inter-state relationships—until the 1990s, but the UN now deems business "an essential partner in achieving the organization's goals."[47] Private firms participate in the UN in a multitude of ways: through engaging in policy dialogue, advocacy, fundraising, and UN operations.[48] Civil society groups have a longer history of involvement at the UN (though this too "dramatically increased" in the 1980s and 1990s), including in the deliberative processes of the UN system, both formally (mainly through the UN Economic and Social Council) and informally (such as when briefing the UN Security Council through the "Arria Formula").[49] More recently, the 2006 Maritime Labor Convention was shaped through a tripartite process involving

[46] "Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies."
[47] Jane Nelson, "'Framework for Business Engagement with the United Nations'" (Global Compact Office, United Nations, September 2008), https://www.un.org/millenniumgoals/2008highlevel/pdf/background/UN_Business%20Framework.pdf.
[48] Benedicte Bull, Morten Bøås, and Desmond McNeill, "Private Sector Influence in the Multilateral System: A Changing Structure of World Governance?," Global Governance 10, no. 4 (2004): 481–98.
[49] "Civil Society," United Nations, October 23, 2014, https://www.un.org/en/sections/resources-different-audiences/civil-society/index.html; "UN System and Civil Society - an Inventory and Analysis of Practices" (Background Paper for the Secretary-General's Panel of Eminent Persons on United Nations Relations with Civil Society, May 2003), https://www.globalpolicy.org/component/content/article/226-initiatives/32330-un-system-and-civil-society.html.

governments, seafarers' trade unions, and shipowners' associations.[50] Both shipping firms and unions are involved as well in enforcement of the treaty provisions.[51]

The inclusivity of international governance arrangements is important to ensuring that AI technologies are developed in a socially beneficial manner. AI has never been the domain of governments alone, and there is a diverse set of established actors involved in its development and impacted by it. Governments have limited expertise to regulate fast-moving AI technologies.[52] Civil society organizations such as Amnesty International have been at the forefront of calls for AI regulation.[53] As noted above, international standards-setting membership organizations such as the IEEE and the ISO have begun developing standards for the use of AI. Even non-profit research organizations rely heavily on private investors.[54] Moreover, most of the targets of AI regulation are multinational corporations. Just a few of the largest American and Chinese tech companies are estimated to account for two-thirds of global spending on AI development.[55] These and other multinational firms hope to avoid having to operate in accordance with an array of nationally determined AI regulations, and will surely seek a major role in crafting the rules they must play by. Otherwise, in the face of regulatory inaction, they may decide the rules themselves.

b. How large are the barriers to entry into the regulated activity?

A related concept are barriers to entry into a regulated activity, which refers to the technical and economic constraints that dictate which actors are able to engage in the activities of the domain, and therefore need to be subject to regulation through a governance arrangement. The size of these barriers influences the design of a regulatory arrangement, and in particular the mechanisms of regulation that are employed. The higher the barriers to entry, all things equal, the smaller the number of actors in the domain. Generally speaking, negotiations to design the governance system and enforcement mechanisms can be more narrowly-focused and can more easily include all relevant actors when barriers to entry into the regulated activity are high. Conversely, it is more challenging to include all of actors who may be impacted by regulation in the negotiation of the global regulatory cooperation system and tougher to design enforcement mechanisms that effectively target all actors when barriers to entry are low.

For example, when the Outer Space Treaty was signed in 1967, space activities were the domain of governments alone. This meant that to serve as an effective constraint on space

---

[50] "INSIGHT: Maritime Labour Convention (MLC) 2006," Skuld, November 3, 2020, https://www.skuld.com/topics/people/mlc-2006/insight-maritime-labour-convention-mlc-2006/.
[51] "Maritime Labour Convention, 2006 (MLC, 2006), Frequently Asked Questions (FAQ), Online Revised Edition, 2012."
[52] Matthew U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," Harvard Journal of Law & Technology 29 (2016 2015): 353.
[53] "The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems," Amnesty International, May 17, 2018, https://www.amnesty.org/en/documents/document/?indexNumber=pol30%2f8447%2f2018&language=en.
[54] Clark and Hadfield, "Regulatory Markets for AI Safety."
[55] Bughin et al., "Artificial Intelligence - The Next Digital Frontier?"

activities, only national governments had to comply. Negotiators of the treaty did not need to consider separately how it would be enforced on non-state actors, since at the time the barriers to entry for developing outer space technology were too high for the private sector to be independently involved. The importance of barriers to entry is even clearer in the case of the Asilomar Conference. Both the research moratorium agreed upon in advance of the Asilomar Conference and the resultant principles of regulating the development of DNA modification technology relied for the most part on self-enforcement by researchers.[56] This was possible because DNA modification research predominantly involved researchers from public institutions (without a profit motive), who all had years of specialized training during which they were acculturated to understand and take seriously the risks of their work and where professional success was linked to reputation and reciprocity within a globally-connected system of scholarly exchange.[57] The barriers to entry were also very high for the Bretton Woods system of fixed exchange rates: only states with their own national currency were engaging in the activity (managing a domestic monetary system) that Bretton Woods aimed to regulate. There were therefore a very limited number of players that had to be included in the design of the Bretton Woods system and that needed to be monitored to ensure compliance with the regulation.

Although there are substantial competitive returns to scale, barriers to entry into the development of AI technologies are generally quite low in many domains. While many blame high barriers to entry for limiting business adoption of AI, in reality barriers to adoption are more related to limited overall digitization, lack of a clear AI strategy (that is, an idea of how and where these technologies will be beneficial to an organization), and the lack of AI talent.[58] For those with a clear idea of how they could put AI to use, there is ample opportunity to adopt many of these technologies. The computer programs used to apply algorithms to new datasets are cheaply or even freely available. While some advanced physical infrastructure of AI—computer servers and chipsets for training and operating machine learning models on—has higher costs, these are falling.[59] The data needed to train machine learning models is one barrier to entry (since larger datasets yield better algorithms) but these data are increasingly publicly available and exist under-leveraged in many organizations.[60] Training for developing and operating AI has in the past been through advanced mathematics and sciences educations, but this information

[56] While federally-funded researchers were required to comply with the 1976 NIH guidelines, most private-sector researchers voluntary complied as well. Judith A. Johnson, "The NIH Recombinant DNA Guidelines: Brief History and Current Status" (Library of Congress, Congressional Research Service, July 7, 1982), https://www.everycrsreport.com/files/19820707_IB82057_a97dc87219d2c9c83df352fd8dd8985de6fde0f0.pdf.
[57] Berg, "Asilomar 1975."
[58] Michael Chui and Sankalp Malhotra, "AI Adoption Advances, but Foundational Barriers Remain" (McKinsey & Company, November 13, 2018), https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain.
[59] Caleb Watney, "Reducing Entry Barriers in the Development and Application of AI," Policy Study (R Street, October 2018), https://2o9ub0417chl2lg6m43em6psi2i-wpengine.netdna-ssl.com/wp-content/uploads/2018/10/Final-No.-153-.pdf.
[60] For example, see this list of publicly available datasets that can be used for training data: Stacey Stanford, Roberto Iriondo, and Pratik Shukla, "Best Public Datasets for Machine Learning and Data Science, Towards AI," Medium, August 28, 2020, https://medium.com/towards-artificial-intelligence/best-datasets-for-machine-learning-data-science-computer-vision-nlp-ai-c9541058cf4f.

is also increasingly freely available. These low barriers to entry have facilitated the global diffusion of AI and its increasingly globally-dispersed development, as noted above. While the advantages of scale may put smaller AI companies at a disadvantage, from a regulatory perspective governments cannot afford to focus exclusively on regulating big technology companies. Network externalities can be fickle: if large subsets of users grow disenchanted or bored with one social media platform, for example, another can easily pop up.

The result of these low barriers to entry is that, unlike regulations around space technology or DNA modification, AI regulations cannot rely on self-enforcement by a group of researchers that has been acculturated to recognize and take seriously the enforcement of rules intended to mitigate the potential harms of AI technologies. Those who work with AI, such as data scientists, are not subject to professional regulation or codes of conduct and currently receive little to no training in the ethics of AI. Even with such training, however, there is the risk that students will not take these courses seriously, as with required professional ethics courses in undergraduate engineering programs.[61] One indication of the above is that ethics codes have so far had limited impact on AI research, development, and application.[62] This could be in part due to the high degree of private sector involvement in developing AI technologies—since in many cases there is a clear conflict of interest between profit motives and self-regulation—although public agencies have also adopted problematic AI-based technologies, such as some facial recognition technologies used by police forces.

c. How important are national security considerations?

Some of the regulatory challenges addressed by the above governance arrangements are more relevant to national security concerns than others. For example, while the safety of medical devices is undeniably important, it is rarely a matter of national security. Therefore, with MDSAP, participating countries can be more comfortable accepting the audit report of a single auditor as meeting the regulatory certification requirements, since there is little risk to national security from doing so. Another example is the Forest Stewardship Council. One reason why governments have been comfortable ceding the responsibility for responsible forest management to non-governmental organizations and the private sector is because it is rarely considered a national security issue. International trade, on the other hand, often involves issues related to national security such as ensuring that a country can maintain a sufficiently strong steel industry as to manufacture armaments in wartime if necessary. The Bretton Woods system helped project American security interests across the non-communist world.63 At the extreme, issues discussed at the UN Security Council nearly always have substantial national security implications. National security concerns have an impact on the willingness of governments to

[61] Anne Colby and William M. Sullivan, "Ethics Teaching in Undergraduate Engineering Education," Journal of Engineering Education 97, no. 3 (2008): 327–38, https://doi.org/10.1002/j.2168-9830.2008.tb00982.x.
[62] Thilo Hagendorff, "The Ethics of AI Ethics -- An Evaluation of Guidelines," February 28, 2019, https://doi.org/10.1007/s11023-020-09517-8.
[63] "The Geopolitics of the United States, Part 1: The Inevitable Empire," Assessment (Stratfor Worldview, July 4, 2016), https://worldview.stratfor.com/article/geopolitics-united-states-part-1-inevitable-empire.

cooperate on standards and monitoring and can also be a powerful source of resistance to regulation (if that is perceived to hamper technologies that promote national security.)

Despite varied levels of government interest and involvement over time, the national security implications of AI technologies have always been clear. The application of early computing to cryptography during the Second World War demonstrated an early military application of proto-AI technology, despite a later focus on applying proto-AI to mathematical problem-solving and chess.64 An early emphasis on preventing the international proliferation of strong cryptographic technologies outside the US—grounded in national security concerns—limited the degree of international cooperation in this area.65 As discussed above, the development of AI-based technologies in weapons systems, military vehicles, and intelligence analysis has led to the increasing identification of AI with national security issues.

d. What is the availability of 'outside options' to avoid regulation?

Certain regulatory challenges are characterized by having stronger or weaker 'outside options'. This means that for some issues, it is relatively easier for a country or corporation to do as well when avoiding compliance with the rules of a governance system, such as by simply not participating in the governance system. For example, there are several outside options to participating in the Forest Stewardship Council: while it is the best-known sustainable management certification system, there are other similar arrangements, and forests and private firms could operate without any certification. Since the FSC is the most well-known certification and enough consumers demand certified wood products, the FSC has succeeded, but 90% of the world's forests are not FSC certified.[66] Similarly, the Asilomar Conference was successful in ensuring DNA modification research was conducted in a safe manner, but it would not have been difficult to avoid its voluntary mechanisms by remaining completely outside the process. It is challenging to achieve meaningful, effective, and inclusive rulemaking by institutions if their members can do just as well by not being in the institution. For example, international cooperation on cryptography standards has been limited by nations' ability to determine their own standards and enact export controls.[67]

In contrast, the outside options for the European Union Single Market are limited: while a country could opt-out from the system of harmonized standards, for example, they would find that their products would not be accepted for imports in many neighbouring markets. Despite stark divides over most aspects of Brexit, over 90% of British voters support the UK remaining in the EU Single Market.[68] Club-type institutions where you get kicked out for non-compliance are

---

[64] "History of Artificial Intelligence," Council of Europe, accessed November 9, 2020, https://www.coe.int/en/web/artificial-intelligence/history-of-ai.
[65] Leung, "Who Will Govern Artificial Intelligence? Learning from the History of Strategic Politics in Emerging Technologies."
[66] "Overview: Forest Stewardship Council, United States" (Forest Stewardship Council, n.d.), https://us.fsc.org/download.fsc-overview.236.htm.
[67] Leung, "Who Will Govern Artificial Intelligence? Learning from the History of Strategic Politics in Emerging Technologies."
[68] John Curtice, "What Do Voters Want from Brexit?," What UK Thinks: EU (NatCen Social Research, November 16, 2016), https://whatukthinks.org/eu/wp-content/uploads/2016/11/Analysis-paper-9-What-do-voters-want-from-Brexit.pdf.

much more effective if there is a real loss for members that are excluded from the institution. This indicates, however, that the relationship is bicausal to some degree. To an extent, the availability of outside options can be impacted by the institutional arrangements. For example, the wide scope of WTO membership, especially among the largest economies, means there are few good outside options to joining it. The 'club of WTO pariahs' is not very attractive.[69] But this differs from Cold War-era GATT, which did not include most communist countries who instead traded mostly amongst themselves.[70] In general, an arrangement that includes a larger share of potential members means that for those members there are fewer other countries that they could engage with if not participating in the arrangement, and hence more limited outside options (although breakaway coalitions are possible). It may be challenging, however, to spark this positive-feedback loop of building meaningful regulations that limit outside options, which then incentivizes countries to join the international arrangement. One mechanism is treaties that take effect only when ratified by a certain number of signatories. The Maritime Labor Convention was initially signed in 2006 but only went into effect in 2012 when at least 30 member countries of the International Labour Organization, representing one-third of the world's gross shipping tonnage, had ratified the treaty.[71]

A challenge for AI regulation is that currently there are strong outside options to global regulatory cooperation. Given the lack of meaningful regulations on AI, it is difficult to argue that those outside existing principles-based agreements are any worse off than signatories. Another challenge is that, unlike the case when importing and exporting bananas, or even automobiles, the 'country of origin' is much harder to determine for intangible AI-based products. This makes it more difficult to verify that an AI product has indeed been developed by a country that is in accordance with some international arrangement (for example related to minimum requirements for training data), since a firm in that country with tougher, harmonized standards could surreptitiously hire AI-related services from a firm in a country with more flexible standards or weaker enforcement.

7. Tradeoffs in the design of global governance and regulatory systems

While the design of a regulatory institution is informed by attributes of its regulatory challenge, there remains a significant degree of flexibility over how the institution is designed. One way of framing the options is through tradeoffs that describe the different ways that regulation in some domain could be approached. Understanding how the above exemplars fit into these tradeoffs sheds light on the opportunities and constraints facing the design of a global regulatory system for AI.

---

[69] "WTO Members and Observers," World Trade Organization, 2020, https://www.wto.org/english/thewto_e/whatis_e/tif_e/org6_e.htm.
[70] Francine McKenzie, "GATT and the Cold War: Accession Debates, Institutional Development, and the Western Alliance, 1947–1959," Journal of Cold War Studies 10, no. 3 (July 1, 2008): 78–109, https://doi.org/10.1162/jcws.2008.10.3.78.
[71] "Milestone Ratifications of Seafarers' Labour Rights Charter."

a. The trilemma of global integration

One of the most useful concepts for grappling with the challenges of designing institutions to address global issues is the trilemma of globalization. Developed separately by Dani Rodrik and by Lawrence Summers in the context of international political economy and public economic management (respectively), this trilemma posits that international policy arrangements can only fully achieve two of three goals: international integration, national sovereignty, and democratic purpose.[72]

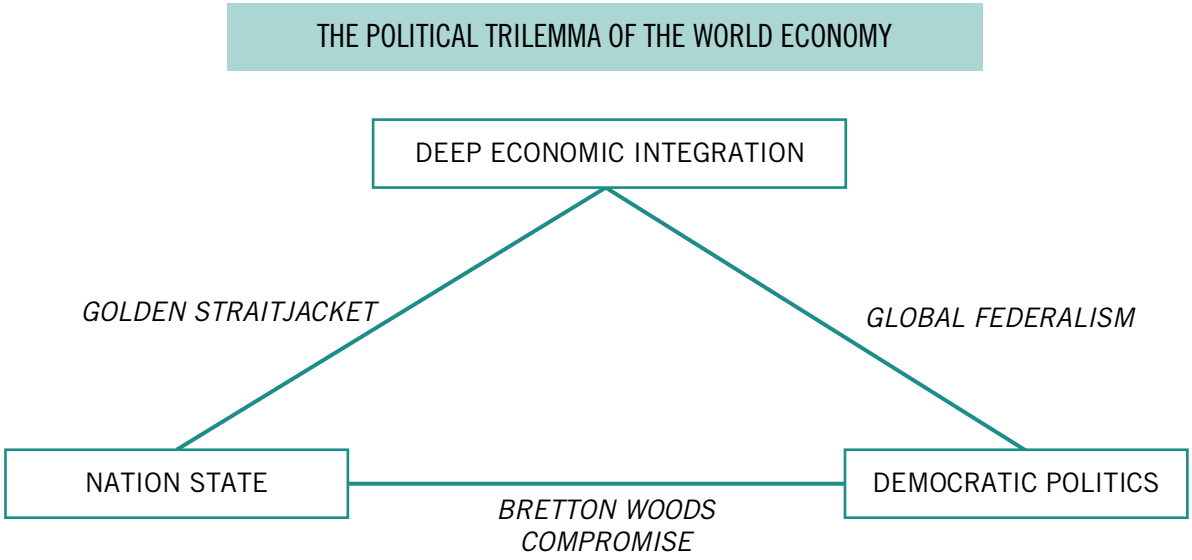
These goals are described both generally, and in the context of AI regulation, as follows:

- International integration: the degree to which countries can easily interact with each other, for example through the flow of goods and services, capital, or people. In the context of AI regulation, this would in the extreme cases translate to uniformity of AI regulations in different jurisdictions, but in other cases could instead involve local regulatory regimes that avoid erecting barriers to the deployment of AI technologies from other locales.
- National sovereignty: the ability of governments to determine their own rules, for example regarding trade and investment restrictions, or capital requirements for banks. For AI, this describes the degree to which governments can establish the regulatory priorities and goals in their jurisdiction, for example regarding the use of facial recognition technologies by local police forces. Several other regulatory issues around AI "go behind the border" including privacy policy, data localization, technology standards, and industrial regulation.[73]
- Democratic purpose: how responsive policies are to the interests of citizens and democratic decision-making, for example whether a government enacts a minimum wage law that has popular support. In the AI regulation context, this would refer to the degree to which AI regulations are as desired by and beneficial to the public at-large. For example, whether data is used in ways consistent with community values.[74]

The consequence of being able to fully achieve only two out of three goals is that there is a choice among three options for an international arrangement on the issue in question. In Rodrik's use of this concept to explain choices in the institutions of international political economy, this choice set is diagrammed as in the following figure:

---

[72] Note that we have utilized "democratic purpose" which aims to combine Summers' concept of "public purpose", with Rodrik's "democratic politics"/"mass political participation". Ch. 9 in Dani Rodrik. The Globalization Paradox. Democracy and the Future of the World Economy. W.W. Norton & Company, 2011; Lawrence H. Summers, "Distinguished Lecture on Economics in Government: Reflections on Managing Global Integration," The Journal of Economic Perspectives 13, no. 2 (1999): 3–18.
[73] Goldfarb and Trefler, "AI and International Trade."
[74] Kevin J. Delaney, "The Robot That Takes Your Job Should Pay Taxes, Says Bill Gates," Quartz, February 17, 2017, https://qz.com/911968/bill-gates-the-robot-that-takes-your-job-should-pay-taxes/.



Each side of this triangle is described as follows:[75]

- Global federalism (limiting the nation state): the "scope of (democratic) politics" is aligned with "the scope of global markets." That is, some sort of supranational governing body sets policies that apply to the residents of all areas that are part of this 'global federation' (note that it need not extend across the entire world). Harmonized standards set by the European Commission regarding the Single Market and applying across the entire EU are the clearest case of this choice.
- Golden straitjacket (limiting democratic politics): the nation state is maintained, but it is "responsive only to the needs of the international economy…at the expense of other domestic objectives." Rodrik's example is the 19th century gold standard, where national governments chose to prioritize the need to maintain global integration (trade and capital flows) through the gold standard over the need for expansionary monetary policy that would benefit their populations.
- Bretton Woods compromise (limiting global integration): a "limited form of globalization" where nation states sacrifice some of their ease-of-interaction for the ability to enact democratically-determined policies. Rodrik's example is the post-war Bretton Woods compromise where international institutions such as the GATT and IMF limited integration by allowing signatory countries to impose capital controls and a substantial degree of trade protectionism (particularly for standards and other non-tariff barriers to trade).

[75] Dani Rodrik, "The Inescapable Trilemma of the World Economy," Dani Rodrik's Weblog (blog), June 27, 2007, https://rodrik.typepad.com/dani_rodriks_weblog/2007/06/the-inescapable.html.

Most international treaties represent a shift towards a form of global federalism. For example, the Outer Space Treaty limits each signatory country's ability to exercise its sovereignty regarding the ability to launch from its soil whatever it would like into space. The Maritime Labour Convention means that countries have to have a minimum degree of labour protections (determined by the Special Tripartite Committee) for shipping vessels that fly their flag and that enter their ports—these countries are no longer free to determine their maritime labour rules as they see fit. In general, it will be easier (but by no means easy) to have countries agree to give up some sovereignty when these countries are similar or have close preferences regarding the issue.

But other arrangements sacrifice some integration to maintain national sovereignty and democratically-determined policies that have a public purpose. International governance of cryptography has involved limited regulatory cooperation and integration. Another example is education. Each country often has very different higher education systems, with differences among the names of degrees, the years of coursework required for a level of degree, accreditation processes, and more. These often make it difficult to transfer a credential from one country to another. However, most governments have determined they would rather keep this policy power than harmonize their post-secondary degree systems—with the exception of the EU, which standardized higher education systems in 1999 in order to promote learning and labour mobility.[76]

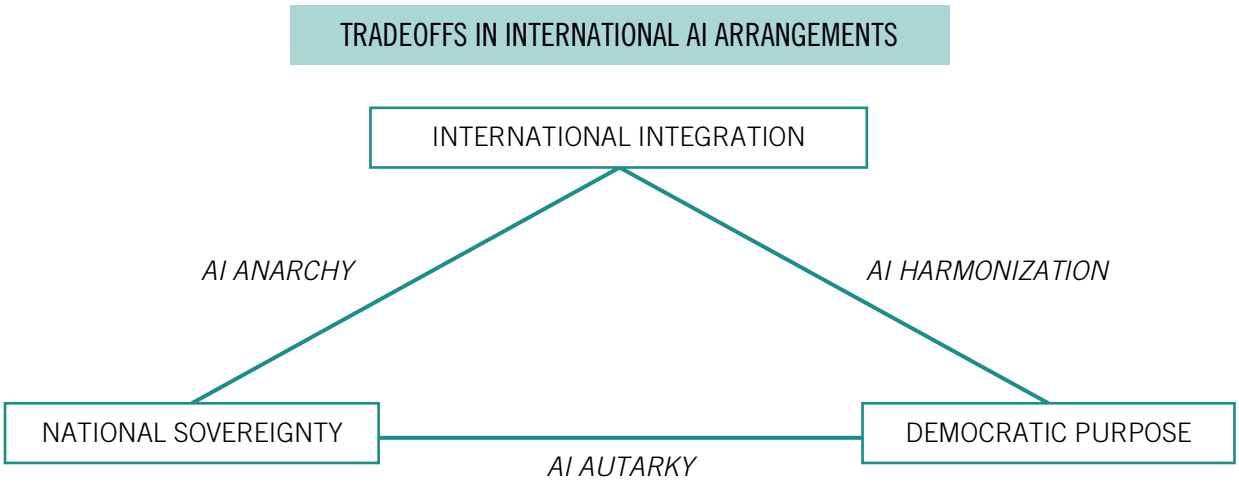
This trilemma of global integration can be applied to describe the tradeoffs involved with different international arrangements in the AI context. The choices would be as follows:

- AI autarky (limiting integration): AI technologies would be very limited in the extent to which they can be jointly developed or adopted from one country to another. For example, it would be very costly for a facial recognition algorithm developed in Korea to be compliant with regulations in exports markets in Sweden, Brazil, and other countries because the relevant regulations would have been crafted independently of each other.
- AI harmonization (limiting sovereignty): this would be a form of 'global federalism' in which regulations on AI apply to all jurisdictions that are part of the arrangement. Note it does not have to be across the entire world—rather, there is a set of universal rules and policies only for the subset of jurisdictions that are participating. For example, if an EU regulatory body set some uniform standards for the definition and size of training and validation sets for medical AI in order to ensure machine learning algorithms replicate well.
- AI anarchy (limiting democratic purpose): Each country would decide to put little regulation on AI, despite there being public support for some sort of restrictions. This could occur in a 'race to the bottom' (for example, loosening previously-required audits

[76] "The Bologna Process and the European Higher Education Area," Text, Education and Training - European Commission, September 21, 2018, https://ec.europa.eu/education/policies/higher-education/bologna-process-and-european-higher-education-area_en.

for algorithmic fairness) whereby each country wants to be the hub of AI development, so no countries put restrictions on AI in order to attract firms to locate there. It could also characterize a situation whereby corporations set their own standards related to AI. In this sort of arrangement, a nation would always resolve conflicts between public desires and the demands of integration towards the latter.

These choices can be diagrammed as follows:



To illustrate these tradeoffs, consider the challenge of dealing with the AI displacement of jobs. There would be no conflict between the competing goals if there is no public will for national governments to enact policies addressing this displacement, because then the national governments can be sovereign, acting with democratic purpose, and maintain integration by using the same AI technologies as elsewhere. However, if public opinion shifted towards preferring some sort of regulatory intervention against AI job displacement, then this national government would be forced to either give up integration (not allowing AI technologies linked with job displacement that are allowed elsewhere) or hope that other jurisdictions adopt the same regulations.

The framing of these tradeoffs illustrates the challenges of past efforts at international cooperation on AI. For the most part, this cooperation has been limited to shallow adoption of principles with little, if any, prospect of enforcement. One example is the International Panel on Artificial Intelligence announced by Canada and France in December 2018 (since renamed the Global Partnership on AI). While this body was intended to build consensus among the G7 countries on limits on AI technologies, for example to protect human rights, it faced opposition from the US.[77] Although the body offered little prospect of enacting binding and enforceable

[77] Tom Simonite, "The World Has a Plan to Rein in AI—but the US Doesn't Like It," Wired, January 6, 2020, https://www.wired.com/story/world-plan-rein-ai-us-doesnt-like/.

regulation on signees, the US still decided it did not want to have externally-imposed constraints on the development of AI by US companies. In essence, the US found this arrangement—where countries sacrificed some sovereignty for integration—unattractive according to its preferences for encouraging AI innovation over safety. As mentioned above, countries that are more similar, especially in their alignment towards AI, are more likely to find meaningful agreement on AI restrictions, as suggested by the recent cooperation between Nordic and Baltic countries.[78]

b. Compliance and depth versus comprehensiveness of membership

A related tradeoff for the design of global governance and regulatory systems is the degree to which the systems force members to comply with regulations, versus the scope of membership for these international arrangements. This tradeoff expands on the tension in the above trilemma between having more or less integration within an international arrangement. It is difficult to get meaningful and effective rulemaking by institutions that have many members with divergent viewpoints. Institutions with many members that have very different (or even competing) interests on an issue will not come to any sort of meaningful agreement on it. In effect, the scope of the arrangement (that is, who is included) determines the purpose and design of institutions: 'who is at the table' helps determine what is 'on the menu'. This tradeoff is true for institutions with member countries as well as those with corporate and NGO members. For example, the Forest Stewardship Council utilizes a complex decision-making process with three membership chambers (Environmental, Social, and Economic) each split into high-income and non-high-income sub-chambers.[79] This structure is designed to maintain flexibility in the balance of voting power between different interests.

In practice, there are several mechanisms that international arrangements use to navigate this tradeoff between deep compliance and comprehensiveness of membership. The most straightforward is to have a very shallow depth of regulations. The Maritime Labour Convention (MLC), for example, sets only *minimum* standards for the living and working conditions on ships. While the MLC covers over 90% of the world's shipping fleet, these minimum standards have been criticized as overly lax and largely discretionary.[80] The Forest Stewardship Council has similarly faced criticism for having less-than-stringent certification processes, which caused Greenpeace to withdraw its membership in 2018.[81]

---

[78] "Ambitious Targets for Artificial Intelligence in the Nordic and Baltic Countries - Nordic Cooperation," accessed November 9, 2020, https://www.norden.org/en/news/ambitious-targets-artificial-intelligence-nordic-and-baltic-countries.
[79] "FSC's Unique Governance Structure" (Forest Stewardship Council, July 2011), https://ic.fsc.org/file-download.principles-and-criteria-v5-web.a-47.pdf.
[80]Marina Fotteler, Olaf Jensen, and Despena Andrioti, "Seafarers' Views on the Impact of the Maritime Labour Convention 2006 on Their Living and Working Conditions: Results from a Pilot Study," International Maritime Health 69 (December 20, 2018): 257–63, https://doi.org/10.5603/IMH.2018.0041.; Paul Bauer, "The Maritime Labour Convention: An Adequate Guarantee of Seafarer Rights, or an Impediment to True Reforms?," Chicago Journal of International Law 8, no. 2 (January 1, 2008), https://chicagounbound.uchicago.edu/cjil/vol8/iss2/12.
[81] "Greenpeace International to Not Renew FSC Membership," Greenpeace International, March 26, 2018, https://www.greenpeace.org/international/press-release/15589/greenpeace-international-to-not-renew-fsc-membership.

International institutions may also have decision-making processes structured so that they limit the actual enforcement of regulations. For example, the former GATT dispute settlement mechanism required rulings to be adopted by consensus, meaning a single country could block a finding of trade liberalization rules being violated (including a finding against themselves).[82] This had the effect of limiting liberalization, especially with regards to more controversial rulings on non-tariff trade barriers. In contrast, after 1995 the WTO made it impossible for the country losing a case to block the adoption of the ruling, which has facilitated greater enforcement of trade liberalization rules. This WTO system provides for sanctions to be set at a more efficient level.[83] Another example is in the difference between the majority voting required to pass non-binding resolutions in the UN General Assembly, but consensus required among the Permanent Five members of the UN Security Council to pass binding UNSC resolutions. Some international treaties, for example for the International Court of Justice, even include provisions allowing signees to opt-out of otherwise-required treaty commitments.[84]

However, since the motivation for creating a global governance system is to craft effective regulatory interventions on an issue, then the above mechanisms can have limited appeal. Compliance is also important for ensuring investor confidence and promoting innovation (for example, through enforcing IP protections). Another approach is therefore to target the other side of this tradeoff: to limit membership in the arrangement only to countries that have closer preferences regarding the key issues. One example of this is the European Single Market which, while quite large, does not include other geographically proximate countries such as Russia or Turkey that likely have very different preferences regarding product standards. The Bretton Woods system of fixed exchange rates included only a limited number of member countries who were closely monitored for compliance to agreed-upon currency valuation.

Any global arrangement to govern or regulate AI must therefore contend with the question of membership. Given current national security trends, an institution that includes the US, China, Russia, and other belligerent nations will look and behave very differently from an organization that includes more aligned countries such as Canada, the EU, Australia, and New Zealand. Some AI researchers suggest that "in the context of the ethical development of artificial intelligence… coordination for a common good should be attempted in smaller groups in which the cost for cooperation is low, and the perceived risk of failure is high."[85]

c. Accountability versus ability to agree

A final tradeoff in the design of international institutions relates to their 'license to operate'—through having a mandate from and being accountable to domestic constituencies—versus their

[82] "Understanding the WTO - Settling Disputes."
[83] Warren F. Schwartz and Alan O. Sykes, "The Economic Structure of Renegotiation and Dispute Resolution in the World Trade Organization," The Journal of Legal Studies 31, no. S1 (2002): S179–204, https://doi.org/10.1086/340406.
[84] Jean Galbraith, "Treaty Options: Towards a Behavioral Understanding of Treaty Design," Virginia Journal of International Law 53, no. 2 (January 1, 2013): 309.
[85] Travis LaCroix and Aydin Mohseni, "The Tragedy of the AI Commons," ArXiv:2006.05203 [Cs], June 9, 2020, http://arxiv.org/abs/2006.05203.

ability to find agreement among members on the global stage. This is not always an issue, but does emerge when the preferences of each side's constituencies diverge (which is typically an assumption of the trilemma above—otherwise the Harmonization of AI option would not be problematic). If the preferences of the constituencies that each side is responsible to do diverge, then global institutions that are more responsive to their constituencies will have greater difficulty coming to agreements on aspects of certain regulatory interventions (or will simply 'agree to disagree'). For example, non-democracies will have an easier time finding agreement at the WTO than democracies, because the former can more easily agree to policies that (temporarily) hurt their domestic populations. Similarly, the greater attention European publics pay to policies of the European Single Market, the harder it will be for national representatives to move from their domestic standards towards harmonization. The Bretton Woods system had high accountability, as demonstrated by its eventual demise when the US suspended its gold standard in part as a response to public pressure against "the attacks of international money speculators" on the dollar.[86]

ISO experts likely have an easier time operating by consensus due to their distance from mechanisms of democratic accountability. ISO members can be governmental or public entities (80% of members) or may be public or private organizations (that often have some sort of institutional link with their national government).[87] For example, the Standards Council of Canada, a crown corporation reporting to Parliament through the Minister of Industry, manages Canadian participation at the ISO. In contrast, the Colombian Institute of Technical Standards and Certification, the ISO member for Colombia, is a non-profit private organization with open membership from the government and private sector. Despite some links between ISO members and government institutions, however, the typical degree of delegation by governments to the ISO for standards-setting, and the general low salience of many standardization issues may limit accountability. Indeed, for standards-setting "acting through seemingly private bodies is an old trick of governments that seek to evade constitutional constraints and public accountability."[88]

For global AI governance, while it is important that international regulatory bodies are in the end accountable to those they represent, they would reach more meaningful agreements if there are mechanisms that afford them latitude in deliberation and decision-making. For example, having national delegates to an international AI regulatory body selected for fixed terms (especially of a relatively long duration) would insulate them from day-to-day political trends, and make it easier to find agreement and thus achieve integration.

---

[86] "Nixon and the End of the Bretton Woods System, 1971–1973," Office of the Historian, , Foreign Service Institute, United States Department of State, accessed November 9, 2020, https://history.state.gov/milestones/1969-1976/nixon-shock.

[87] Jeanne Dupendant, "International Regulatory Co-Operation and International Organisations: The Case of the International Organization for Standardization (ISO)" (OECD and ISO, 2016), https://www.oecd.org/gov/international-regulatory-co-operation-9789264244047-en.htm.

[88] Eyal Benvenisti, "Towards a Typology of Informal International Lawmaking Mechanisms and Their Distinct Accountability Gaps," in Informal International Lawmaking (Oxford University Press, 2013), https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199658589.001.0001/acprof-9780199658589-chapter-14.

8. Conclusions for the global governance and regulation of AI

The above discussions outline how eleven exemplars of global regulatory systems have been shaped by their regulatory challenge and the tradeoffs in designing institutions of global governance. These have several implications for thinking about the design of an AI global governance and regulation system.

a. Success is not guaranteed when crafting global cooperation on regulation

The focus on exemplars of governance arrangements that successfully addressed a regulatory challenge is helpful for drawing lessons for what may work for AI governance. However, it may also result in some 'survivorship bias'—that is, we know less about why a global governance initiative may fail to implement meaningful regulation. It is important to contend with the possibility of there being little meaningful international cooperation on AI governance.

The limited amount of international cooperation related to cryptographic technology is illustrative of this possibility. The lack of meaningful global cooperation on regulating cryptographic technologies is in part because nations have pursued cryptography primarily as a tool for national security.[89] Technological development was mainly the domain of government for the first few decades of cryptography, and although it has since shifted to predominantly private sector involvement these businesses have largely aligned their interests with national governments. As a result, the global "market is fragmented amongst several non-interoperable suites of cryptographic algorithms."[90] While nations would benefit from some degree of cooperation that facilitated greater technological development, they also have strong outside options to deploy cryptography primarily for national security applications. Existing international cooperation is limited to the 1997 principles by the OECD limited to member countries, and some rapporteur reports adopted by UNESCO that propose to facilitate dialogue.[91] Without real action, it is very possible that AI international governance will continue to be limited to expressions of principles by international organizations.

b. Multilevel international arrangements offer an attractive framework

The discussions above demonstrate that it is possible to have multilevel international arrangements that include various membership roles, depths of commitment, and selectivity of regulation. For example, the WTO has minimum tariff rates set at the most favoured nation (MFN) bound rates, but also has bi- and multi-lateral agreements among its members that prescribe lower effective tariff rates and closer harmonization of standards (among other features). This serves as a template for a flexible mechanism for international AI regulatory cooperation, since it could enable different degrees of integration (e.g. harmonization of AI standards) and a menu of regulatory actions that could be taken in collaboration with other countries. This responds to calls for a global AI governance system "flexible enough to

[89] Leung, "Who Will Govern Artificial Intelligence? Learning from the History of Strategic Politics in Emerging Technologies."
[90] Leung.
[91] "OECD Guidelines for Cryptography Policy - OECD."; Schulz and van Hoboken, Human Rights and Encryption.

accommodate cultural differences and bridge gaps across different national legal systems."[92] It could also achieve a level of 'minimalist integration' based on cooperation that does not require agreement on standards for all areas of AI and can "reach agreement on practical issues despite disagreement on more abstract values or principles."[93]

In the case of further divergence among national preferences for AI regulation, a large multi-level regulatory arrangement could still preserve some degree of minimum standards, even in the case of global fragmentation into different AI regulatory blocs (as, for example, is predicted with countries choosing firms that provide 5G infrastructure that are backed by China or the US). A multi-level arrangement could also open the door to a convergence towards the most stringent regulations through the so-called "Brussels effect."[94] Flexible integration that can gradually become closer may eventually enable better regulation. Research on competition policy suggests that high-quality regulation will be produced by a centralized supra-national regulator that "will be tilted towards more independence" than a national regulator and thus better able to avoid capture and maximize consumer surplus.[95]

c. A successful global AI regulatory system requires real benefits and enforcement

Given the need to sacrifice either some national sovereignty or democratic purpose in order to achieve international integration, any global AI regulatory arrangement must have tangible benefits to attract members. There are several potential benefits of membership for AI. For example, many AI professionals would only work for members and member companies would not be able to collaborate with non-members in joint ventures. In addition, universities and other public research institutions could restrict partnerships to members and governments could limit procurement contracts to members. If the cooperation has sufficiently robust standards, there would also be public trust associated with membership.

A successful global regulatory arrangement needs some sort of enforcement mechanism. Given the tradeoff described above between compliance and comprehensiveness, these mechanisms may be weaker in some cases, in order to achieve minimum standards. In an arrangement with more limited membership, more robust standards could be achieved. There are several options for enforcement of these standards. For example, a certain company or products could be blacklisted, or subject to fines. AI training opportunities for nationals of a country, or data access, could be restricted. There could also be sanctions placed on member countries or companies deemed liable for some breach.

---

[92] Urs Gasser and Virgilio A. F. Almeida, "A Layered Model for AI Governance," IEEE Internet Computing 21, no. 6 (November 2017), https://doi.org/10.1109/MIC.2017.4180835.
[93] Seán S. ÓhÉigeartaigh et al., "Overcoming Barriers to Cross-Cultural Cooperation in AI Ethics and Governance," Philosophy & Technology 33, no. 4 (December 1, 2020): 571–93, https://doi.org/10.1007/s13347-020-00402-x.
[94] The Brussels Effect: How the European Union Rules the World (Oxford, New York: Oxford University Press, 2020).
[95] Germán Gutiérrez and Thomas Philippon, "How European Markets Became Free: A Study of Institutional Drift" (National Bureau of Economic Research, June 11, 2018), https://doi.org/10.3386/w24700.

9. Closing

Promising AI technologies require an effective global regulation and governance system in order to maximize the social welfare of these innovations. This report has described how we can learn from the experiences of past international regulatory cooperation when crafting a global regulatory system for AI. As with the post-WWII Bretton Woods system, there is now an impetus for real action on regulating the harms of AI. It is crucial that the global regulatory system designed to accomplish this goal is set-up to flourish.

SCHWARTZ REISMAN INSTITUTE FOR TECHNOLOGY AND SOCIETY

The Schwartz Reisman Institute aims to deepen our knowledge of technologies, societies, and what it means to be human by integrating research across traditional boundaries and building human-centred solutions that really make a difference. We want to make sure powerful technologies truly make the world a better place—for everyone. Comprising diverse areas of inquiry, from machine learning, computer engineering, epistemology, systems theory, and ethics to legal design, systems of governance, and human rights, our research agenda and solutions stream cross traditional boundaries and are fundamentally inspired by a commitment to reinventing from the ground up.