OXGS Research Report

Navigating Geopolitics in AI Governance

April 2024

Oxford
Global
Society

Authors:
Jufang Wang, Claire Milne, Jess Zichen Hu, Furqan Khan

About the authors

Dr Jufang Wang is the Deputy Director of Oxford Global Society (OXGS), where she coordinates the work of the Digital Technologies Cluster. Her research focuses on digital technology governance and online public opinion.

Claire Milne is a former Visiting Senior Fellow and Guest Teacher at the London School of Economics (LSE), and a Senior Advisor to the OXGS. She has been an independent consultant since 1989, providing ICT policy and regulatory advice in dozens of countries. She has served on several public bodies in the UK. In 2015 she was awarded an MBE for services to the telecommunications sector.

Jess Zichen Hu is a doctoral researcher at the Media and Communications Department of LSE. Her research focuses on digital technologies and governance. She is also a research assistant at the Oxford Global Society.

Furqan Khan is a researcher at the Islamabad Policy Research Institute (IPRI, Pakistan) and a research assistant at the Oxford Global Society. His research centers on International Relations, with geopolitics being a special focus.

Table of Contents

Acknowledgements

Executive summary

1. The three major AI powers (i.e., the EU, the US, and China) employ different digital regulatory approaches that differ in their emphases on rights-driven, market-driven, and state-driven objectives. While sharing a common governance goal of achieving a delicate balance between innovation and AI safety, these approaches reflect underlying values and priorities in governing AI: prioritizing human rights, innovation, or social stability, or a combination of them.

2. Despite differences in regulatory approaches, countries/regions have many commonalities in the values and principles underlying AI governance, such as human-centered values. These shared values and principles are often overlooked and should be emphasised to facilitate global collaboration on AI development and governance.

3. Differences between values-based and hardware-based governance modes demonstrate tensions in AI governance. While values-based governance can facilitate cooperation between like-minded countries, it may also lead to conflict in a global arena with differing values. In contrast, hardware-based governance may be used to reduce AI misuse and enhance safety, but it can also be deployed to seek control over other countries' AI development and thus exacerbate geopolitical rivalry.

4. Geopolitics is a major factor that shapes nations' (or regions') AI governance strategies and practices. There is a tendency to frame geopolitical competition around AI in terms of a US-China two-way rivalry. However, geopolitical developments need to be understood as including "middle powers" as well because of their aspirations to play a role in AI development and governance.

5. Computing power is being used as an AI governance tool (for example, in the EU AI Act) and for "leverage" in geopolitical competition. The US is resorting to measures such

as export controls on semiconductor chips and economic protectionism policy to consolidate its dominance in AI technologies and to contain China's AI development. This intensifies technology rivalry, disrupts AI supply chains, and hinders global cooperation on AI governance.

6. Geopolitical tensions and the nature of advanced AI technologies (e.g., the training of Large Language Models (LLMs) requiring large quantities of computing power, massive data, and complex algorithms) are contributing to the concentration of power in AI and to an expanding divide between AI technology haves and have-nots. The Global South risks being left behind in both AI development and global AI governance.

7. Experts disagree on the potential of various forms of global AI governance frameworks, but there is increasing agreement on the necessity for a centralized global framework to tackle what some see as existential risks of AI. For non-existential risks, experts are more likely to prefer a distributed framework reflecting diverging values across countries/regions and respecting diversity acorss countries/regions.

8. With respect to claimed existential risks of AI, there are signs of increasing global cooperation, and international mechanisms similar to the International Atomic Energy Agency (IAEA) have been suggested as reference points. However, pessimists warn that countries or regions may only cooperate after a disaster, as in the case of chemical and nuclear weapons.

9. Given the many shared values and principles of AI governance among countries or regions, there may be room for a global AI framework that is not limited to addressing existential AI risks. The institutional arrangement of any global AI governance framework is uncertain due to power relations and varying views on the right role for the UN. Many experts favour a hybrid (soft and hard) AI governance approach.

List of Acronyms

AI	Artificial Intelligence
AGI	Artificial General Intelligence
ASML	Advanced Semiconductor Materials Lithography (Holding, Netherlands)
BRI	Belt and Road Initiative
BRICS Plus	Brazil-Russia-India-China-South Africa Plus
BIS	Bureau of Industry and Security (US)
CAC	Cyberspace Administration of China
DSA	Digital Services Act (EU)
EU	European Union
FTC	Federal Trade Commission (US)
GAIGI	Global AI Governance Initiative (China)
GPAI	Global Partnership on Artificial Intelligence
GPU	Graphics Processing Units
GPT-4	Generative Pre-trained Transformer 4
IEEE	Institute of Electrical and Electronics Engineers
IAEA	International Atomic Energy Agency
ICAO	International Civil Aviation Organization
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
LLMs	Large Language Models
LAWS	Lethal Autonomous Weapon Systems
MIT	Massachusetts Institute of Technology
NIST	National Institute for Standards and Technology (US)
OECD	Organisation for Economic Co-operation and Development
OXGS	Oxford Global Society
RMB	Renminbi (China)
SMIC	Semiconductor Manufacturing International Corporation (China)
SMEs	Small and Midsize Enterprises

TSMC   Taiwan Semiconductor Manufacturing Company Limited

TTC   Trade and Technology Council (EU-US)

UN   United Nations

UNESCO   United Nations Educational, Scientific and Cultural Organization

UK   United Kingdom

US   United States

WHO   World Health Organization

1. Introduction

Amidst breakthroughs in AI technologies (such as generative AI), there is a growing impetus to govern this transformative technology to better harness its potential and address the substantial risks it poses. At the national/regional level, the European Union (EU) AI Act, the first comprehensive legal framework on AI worldwide has been agreed.[1] In China, an interim regulation on generative AI took effect in August 2023, making China the first country to pass a regulation on generative AI services.[2] Even in the United States (US), which has been reluctant to regulate digital media platforms or introduce a comprehensive national data protection law, AI has emerged as a robust policy issue for Congress and the Biden Administration.[3]

At the multilateral level, multiple moves towards governing AI have been witnessed in recent years. For example, the OECD has published its values-based AI Principles (in 2019) and released its "Framework for the Classification of AI Systems" (in 2022). It has aso established the "AI Incidents Monitor" to document AI incidents (in 2023), and

---

[1] European Parliament (13 March 2024). *Artificial Intelligence Act: MEPs adopt landmark law*. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law (The Act will become law when published in the Official Journal of the European Commission).
[2] Wu, Y. (27 July 2023). How to Interpret China's First Effort to Regulate Generative AI Measures. *China Briefing*. https://www.china-briefing.com/news/how-to-interpret-chinas-first-effort-to-regulate-generative-ai-measures/
[3] Covington (20 October 2023). U.S. Artificial Intelligence Policy: Legislative and Regulatory Developments. https://www.cov.com/en/news-and-insights/insights/2023/10/us-artificial-intelligence-policy-legislative-and-regulatory-developments

updated its definition of AI (in 2023).[4] UNESCO produced the "Recommendations on the Ethics of AI" in 2021, the first global standards in AI ethics, to ensure that AI developments align with the protection of human rights and human dignity.[5] The UN Secretary-General's high-level AI Advisory Body, comprising 32 experts in relevant disciplines from around the world, launched its *Interim Report: Governing AI for Humanity* in December 2023, calling for closer alignment between international norms and how AI is developed and rolled out.[6] As the report pointed out, there is "no shortage of guides, frameworks, and principles" on AI governance drafted by various bodies (including the private sector and civil society).[7]

However, there is a substantial gap between the rhetoric about governing AI and its implementation on both national and international levels. One major factor hindering the effective implementation of AI governance is geopolitical relations, as these may pressure countries to prioritize national competitiveness over safety concerns about AI systems and discourage cooperation among countries and regions regarding AI governance. Other factors include different emphases on the values underlying regulatory approaches to AI governance and the huge and expanding divide between the haves and have-nots in AI technology capabilities.

In this report, we examine the efforts of major AI powers (the US, the EU, China) towards governing AI and the values underlying them, how geopolitics influences AI governance and its implementation, whether a global AI governance framework is feasible and, if so, what area a framework should focus on and what form it should take. We do not delve into the technical dimensions and processes of AI governance (such as how to audit AI

---

[4] OECD (March 2024). Explanatory memorandum on the updated OECD definition of an AI system. OECD ARTIFICIAL INTELLIGENCE PAPERS, No. 8. https://doi.org/10.1787/dee339a8-en

[5] UNESCO (2021). Recommendation on the ethics of artificial intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000380455

[6] UN Secretary-General's high-level AI advisory Body (December 2023). Interim Report: Governing AI for Humanity. https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf

[7] Ibid., p.12.

algorithms or how to ensure AI systems are explainable and accountable). Specifically, we try to answer the following questions:

- *What are the main differences and commonalities among major AI powers regarding their approaches to AI governance? What are the values or principles underlying the approaches?*
- *What is the role of geopolitics in shaping national/international AI governance?*
- *Given geopolitical competition around AI, is a centralized global AI governance framework possible or necessary? If so, what should it look like (focusing on what areas, where should institutional power lie, etc.)?*

These were the main questions raised and discussed in an online seminar[8] held by the Oxford Global Society in January 2024 involving experts from the US, EU, the UK and China. This report draws on the seminar discussion as the primary source (the seminar is referred to as "OXGS seminar" and speakers at the seminar are referred to as "our panelist(s)"), supplemented by insight drawn from a wide range of literature including academic books, journal articles, policy papers, and media reports. As an independent research report on the interactions of AI governance and geopolitics, our analysis reflects but also goes beyond the seminar discussion.

To structure the report, we first explore what the term "AI governance" means, since it involves various layers and multiple actors and means different things to different actors. This is followed by an analysis of the differences and commonalities in regulatory approaches and the underlying values of the major AI powers and a discussion of two different modes of governance: values-based and hardware-based. Next, we examine the role of geopolitics in AI governance, discussing how the US has resorted to computing power as "leverage" in geopolitical competition and how a concentration of power in AI can be seen as entrenching a divide in AI technology capabilities and governance. Finally,

---

[8] Oxford Global Society (8 Jan 2024). Online seminar: Navigating Geopolitics in AI Governance Implementation (speaker profiles and recording links included). https://oxgs.org/event/navigating-geopolitics-in-ai-governance-implementation/

we consider whether a centralized global AI governance framework is possible or necessary and the forms it could take.

2. "AI governance": Definition and the multi-layer model

The term "AI governance" is understood to cover a variety of approaches to ensuring the research, development, and use of AI systems are responsible, ethical and safe. In this report, AI governance approaches include relevant laws, regulations, rules, and standards. While this term is widely used both within and outside companies and other organizations, the focus here is on AI systems governance practiced by external bodies such as governments, industry alliances, and international organizations.

The use of the term "AI governance" is not without debate. Our panelist Milton Mueller, Professor at Georgia Institute of Technology, questioned the global hype over "AI governance". In his words, what we call "AI" is simply:

> *The latest manifestation of computing...the latest development of increasingly powerful information processing capabilities, increasingly sophisticated software applications, faster networks and larger and larger stores of digital data….*

He therefore argued that any attempt to govern AI would entail governing "the entire digital ecosystem", including the governance of data, clouds, semi-conductors, content moderation of social media, intellectual property rights, cyber security, etc.

We, like other panelists of the OXGS seminar, agree that AI governance implies the governance of the whole digital ecosystem. However, it is also considered that the governance of AI differs from that of previous digital technologies because AI systems, using massive training datasets and "black box" machine learning algorithms, can lead to particularly unexpected risks. For example, unbalanced training data may lead to systematic and entrenched discrimination within an AI system. This unpredictability

feature is reflected in the recently amended OECD definition of AI (both the US's AI Risk Management Framework and the EU's AI Act align with this OECD definition):[9]

> An AI system is a machine-based system that, **for explicit or implicit objectives**, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Compared to an earlier definition, the modified version omitted the "objectives defined by humans" and changed to "explicit or implicit objectives", recognizing the capacity of AI systems to learn new objectives. Whether AI causes "existential risks" is a much debated issue among scientists, researchers, and industry experts.[10] Along with others, we here include in the "existential" category extreme or "societal-scale" risks which fall short of the extinction of human beings. We assume that AI systems such as Lethal Autonomous Weapons Systems (LAWS) and currently hypothetical Artificial General Intelligence (or AGI) and superintelligent AI[11] can bring substantial risks, especially when AI systems become competent to pursue goals misaligned with humans'.[12]

AI governance thus involves multiple layers. This is illustrated by the "multi-layer model" developed by Jovan Kurbalija from DiploFoundation (as illustrated in Figure 2.1), which we use as a reference point in discussing AI governance. This model includes AI governance layers from hardware (or computing power), data (or source for AI), algorithm (or AI capabilities), through to applications (or end-user layer). This is a helpful way to

---

[9] OCED. (nd). AI terms and concepts. https://oecd.ai/en/ai-principles

[10] See for example: Heaven, W. D. (19 June 2023). How existential risk became the biggest meme in AI. *MIT Technology Review*. https://www.technologyreview.com/2023/06/19/1075140/how-existential-risk-became-biggest-meme-in-ai

[11] AGI is also known as strong AI, full AI, or human-level AI, can accomplish any intellectual tasks that human beings can perform. It contrasts with "narrow AI", which performs a specific task. Superintelligent AI refers to an agent that possesses intelligence far surpassing that of humans.

[12] Conn, A. (2015). Benefits & Risks of Artificial Intelligence. *Future of Life Institute*. https://futureoflife.org/ai/benefits-risks-of-artificial-intelligence/

envisage AI governance which involves issues like export controls and sanctions (at the hardware layer), data protection and privacy (at the data layer), transparency and explainability (at the algorithm layer), and consumer protection and antitrust regulation (at the end-user layer). AI governance also involves a myriad of actors such as the private sector (microchip companies, major AI companies like OpenAI, Microsoft, and Google), governments, international organizations, and civil society and consumer organizations. In this report, we focus on governance efforts and initiatives of governments and international organizations.
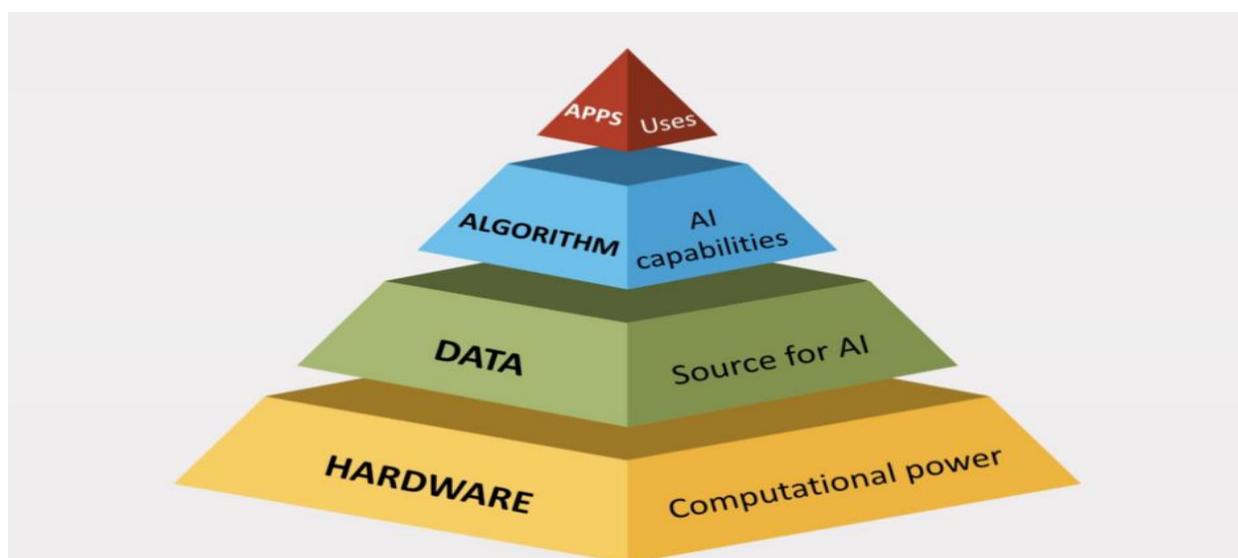
Figure 2.1 Source: DiploFoundation

3. Diverging regulatory models and underlying values of major AI powers

This section investigates the approaches and underlying values of AI governance by the three major global AI powers: the EU, the US and China, [13] which represent AI regulatory models with different emphases. We aim to identify both their differences and commonalities. In our view, often-overlooked commonalities hold promise for global

---

[13] While the EU lags the US and China in the development of a strong private AI sector, it is widely regarded as a "regulatory power" in the digital realm.

cooperation in AI governance. We also consider two modes of AI governance: values-based and hardware-based governance.

3.1 Three major AI governance models

Drawing on Columbia Law School Professor Anu Bradford's recent book *Digital Empires,*[14] we categorize the approaches to AI governance in the EU, the US and China (called "three digital empires" by Bradford) as principally rights-driven, market-driven, and state-driven respectively. These approaches reflect different emphases on values and priorities regarding the governance of AI. This differentiation is not absolute and, in practice, each model has a shared governance goal of balancing innovation and AI safety. All three models are still being developed and implemented.

EU's rights-driven model: Risk-based and human-centric

In the EU, we have seen a "tsunami"[15] of regulations or proposed regulations relevant to digital technologies including AI in recent years. Among them, the Digital Services Act (DSA), which entered into force in November 2022, includes requirements for transparency on recommendation algorithms and bans on targeted advertising on online platforms that profile children or based on special categories of personal data such as ethnicity, political views or sexual orientation.[16] The EU AI Act, which has attracted wide global attention partly due to the so-called "Brussels effect" (referring to the EU's potentially unilateral power to regulate global markets)[17], preserves a risk-based approach, classifying AI systems into four categories: unacceptable risk (e.g., social scoring and manipulative AI), high-risk (e.g., AI systems used in areas including critical infrastructure, medical, education, and employment), limited risk, and minimal risk.[18] Unacceptable-risk AI systems will be banned; high-risk systems will be subject to strict

---

[14] Bradford, A. (2020). *Digital Empires: The Global Battle to Regulate Technology*. Oxford, UK: Oxford University Press.

[15] As described by our panelist Maxime Ricard, who is the Policy Manager of Allied for Startups.

[16] European Commission (23 February 2024). Questions and answers on the Digital Services Act. https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_2348

[17] Bradford, A. (2019). *The Brussels Effect: How the European Union Rules the World.* Oxford, UK: Oxford University Press.

[18] European Commission (24 January 2024). AI Act: Consolidated Text. https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-Overview_24-01-2024.pdf

compliance requirements such as adequate risk assessment and detailed documentation; and limited-risk systems will need to fulfill transparency requirements (i.e., telling end-users AI technologies are used) and voluntary compliance.[19] It is worth noting that minimal-risk systems (including many currently available applications such as AI-enabled video games and spam filters) will not be regulated under this Act, showing that the EU is aiming to limit its regulatory restrictions to sensitive AI systems. The Act also introduces dedicated rules for general purpose AI models associated with systemic risk that require additional binding obligations related to managing risks and monitoring serious incidents. Companies that fail to comply with the Act can face substantial fines, up to 7% of their global annual revenue.[20] This is one reason why the EU's AI governance approach is often referred to as the binding "hard regulation".[21]

This risk-based approach raises some questions about its effectiveness in practice. One issue is the difficulty in deciding whether an AI system falls into the category of unacceptable risk or high-risk because judging such risks is not always straightforward. For example, if a company can substantiate its claim that its AI system meets one of the exception conditions such as "performing a narrow procedural task", then it may not be classified as high-risk despite dealing with sensitive areas such as education or employment.[22] Our panelist Professor Mueller commented that it will be difficult to decide the risk level of an AI system before its deployment. Our panelist Professor Angela Huyue Zhang, Associate Professor of Law at Hong Kong University, also pointed out that, for most cases, the EU AI Act only requires industry self-assessment, making it more akin to a "deterrence strategy".

Another feature of the EU regulatory approach is its emphasis on the so-called "human-centric" approach, ensuring AI works for people and protects fundamental human rights

---

[19] Ibid.

[20] Chee, F. Y. (2021). EU set to ratchet up AI fines to 6% of turnover - EU document. *Reuters*. https://www.reuters.com/technology/eu-set-ratchet-up-ai-fines-6-turnover-eu-document-2021-04-20/

[21] Csernatoni, R. (30 January 2024). How the EU Can Navigate the Geopolitics of AI. *Carnegie Europe*. https://carnegieeurope.eu/strategiceurope/91503

[22] European Commission (24 January 2024). AI Act: Consolidated Text. https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-Overview_24-01-2024.pdf

such as dignity, equality and justice.[23] According to the definition from the EU-US Trade and Technology Council (TTC), "human-centric AI" encourages the empowerment of humans in the design and use of AI systems, and is designed as "tools to serve people with the ultimate aim of increasing human and environmental well-being with respect for the rule of law, human rights, democratic values and sustainable development".[24] Our panelist Dr Gry Hasselbalch, a Senior Key Expert on AI Ethics in the European Commission's International Outreach for a Human-Centric Approach to Artificial Intelligence (InTouch.AI.eu), noted that AI governance is now transitioning to debates on existential risks, making the EU's emphasis on humanity and the essence of being human more relevant.

The EU emphasizes that it advocates a regulatory framework that safeguards ethical and societal values and promotes technological progress and innovation. This is exemplified in the Digital Europe Programme (2021-2027) that aims to strengthen the EU's technological leadership.[25] Another action to boost the competitiveness of European AI companies is the European Commission's innovation package to support European startups and SMEs, including making Europe's AI-dedicated supercomputers available to a large number of public and private users, including innovative European AI startups, to train their LLMs.[26]

US's market-driven model: Pro-innovation and limited government intervention

Compared to the EU's formal regulatory model, the US so far has adopted a light-touch approach to governing AI. In 2020, the US Congress passed the National AI Initiative Act

---

[23] European Commission. (7 June 2022). International outreach for human-centric artificial intelligence initiative. https://digital-strategy.ec.europa.eu/en/policies/international-outreach-ai
See also key values on which the Union is founded in Article 2 of the Treaty of European Union: https://eur-lex.europa.eu/resource.html?uri=cellar:2bf140bf-a3f8-4ab2-b506-fd71826e6da6.0023.02/DOC_1&format=PDF
[24] European Commission (1 May 2023). EU-U.S. Terminology and Taxonomy for Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence, p. 9.
See also Hasselbalch, G. (2021). *Data ethics of power: a human approach in the big data and AI era*. Cheltenham, UK: Edward Elgar Publishing.
[25] European Commission (2021). The Digital Europe Programme. https://digital-strategy.ec.europa.eu/en/activities/digital-programme
[26] European Commission (2024). European approach to artificial intelligence. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

(NAIIA)[27], which focuses on ensuring "continued U.S. leadership in AI research and development" and requires developing a voluntary risk management framework.[28] In January 2023, the National Institute for Standards and Technology (NIST) released its AI Risk Management Framework (AI RMF 1.0), which states that it is intended to be "voluntary" and non-sector-specific, and to provide "flexibility" to organizations of all sizes to implement.[29] NIST also plays a role in evaluating and publicly reporting on the accuracy and fairness of facial recognition algorithms through its ongoing Face Recognition Vendor Test program.[30] A notable action in US AI governance is the AI Executive Order issued by President Joe Biden in October 2023.[31] Measures taken by US federal agencies following this Order include: compelling developers of the most powerful AI systems to report vital information, especially AI safety test results, to the Department of Commerce; and proposing the requirement that U.S. cloud companies must report foreign users that might use their services to conduct AI training (referred to "know your customer" rule).[32] While hailed as a "landmark" action on AI governance, the Executive Order, as argued by our panelist Professor Zhang, is still quite "limited" in its scope and enforcement measures since the major mandatory provision for tech companies is disclosure requirements.

The US's market-driven regulatory approach reflects a policy view shaped considerably by economic liberalism that emphasises minimal state intervention and the primacy of free markets. As Anu Bradford notes, the US model encourages the government "to step aside to maximize the private sector's unfettered innovative zeal—except when it comes

---

[27] United States House Committee on Science, Space, and Technology (2020). H.R.6216 - National Artificial Intelligence Initiative Act of 2020116th Congress (2019-2020). https://www.congress.gov/bill/116th-congress/house-bill/6216

[28] Park, L. (2022). National Artificial Intelligence Initiative. United States Patent and Trademark Office. https://www.uspto.gov/sites/default/files/documents/National-Artificial-Intelligence-Initiative-Overview.pdf

[29] National Institute of Standards and Technology (January 2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

[30] National Institute of Standards and Technology (30 November 2020). Face Recognition Vendor Test (FRVT). https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt

[31] White House (30 October 2023). FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

[32] White House (29 January 2024). Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden's Landmark Executive Order. https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/

to protecting national security".[33] However, in recent years, this approach has been criticized for overlooking ethical considerations and societal impacts and as leading to the concentration of power within a few tech giants.[34] There are many proposed bills on mitigating AI risks (which are unlikely to become law) in the US Congress, demonstrating growing concerns among American lawmakers and the public about the risks associated with AI. Thus far, however, the US is mainly relying on existing laws to regulate AI applications. Several federal agencies, including the Federal Trade Commission (FTC) and Consumer Financial Protection Bureau (CFPB), released a joint statement on 25 April 2023, reiterating their enforcement powers and actions already taken against AI and automated systems.[35] This approach is supported by leading scholars in the US. For example, a committee of MIT leaders and scholars released a set of policy papers on the governance of AI which advocate extending current regulatory and liability approaches to oversee AI and also recommended new arrangements regarding auditing new AI tools.[36]

China's state-driven model: government-led, pro-innovation and pro-stability

As some scholars of China's technology regulation observe, China's approach to AI governance is characterised by strong state-led direction combined with active engagement from the private sector.[37] This public-private partnership aims to create a synergistic environment where government support and oversight go hand in hand with market-driven innovation, ensuring that AI development is both ambitious and aligned with national policies and objectives.

---

[33] Bradford, A. (2020). *Digital Empires: The Global Battle to Regulate Technology*. Oxford, UK: Oxford University Press. p.24.

[34] Ibid.

[35] United States Federal Trade Commission. (25 April 2023). Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems. https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf

[36] Dizikes, P. (11 December 2023). MIT group releases white papers on governance of AI. *MIT News*. https://news.mit.edu/2023/mit-group-releases-white-papers-governance-ai-1211

[37] See Professor Angela Huyue Zhang's forthcoming paper: Promise and perils of Chinese AI regulations. Preprint available 12 February 2024 at http://dx.doi.org/10.2139/ssrn.4708676. See also (1) Lee, K. F. (2022). *AI Superpowers: China, Silicon Valley, and the New World Order*. New York, NY: Harper Business, and (2) Liu, T., Yang, X., & Zheng, Y. (2020). Understanding the evolution of public–private partnerships in Chinese e-government: four stages of development, *Asia Pacific Journal of Public Administration, 42*(4), 222-247. https://doi.org/10.1080/23276665.2020.1821726/

In recent years, the Cyberspace Administration of China (CAC), together with other government agencies, has promulgated regulations relevant to AI. The *Regulation on Algorithmic Recommendation of Internet Information Services,* which took effect from March 2022, stipulates that algorithmic recommendation should: (1) ensure algorithms do not lead users (especially minors) towards bad outcomes like addiction and excessive consumption; (2) treat users in a non-discriminatory way; and (3) give users a choice to turn off personalized recommendation.[38] Another relevant regulation is on deep synthesis (or "deepfake") technology, which requires deepfake services like face swapping to be immediately recognizable if they may mislead users.[39] The most recent regulation is the *Interim Administrative Measures for the Management of Generative AI Services* (referred to as "China's generative AI regulation", which took effect from 15 August 2023).[40] While emphasizing principles such as socialist values, non-discrimination, privacy protection, and transparency (labeling content generated by AI), this regulation explicitly encourages innovation in developing generative AI algorithms and chips, engagement in international exchange and cooperation, and participation in the formulation of international rules related to generative AI.[41] Unlike the EU AI Act, China's regulations relevant to AI do not specify any fines or provide for issuing very small fines (e.g., under 100k RMB or around 14k US dollars) for non-compliance. However, the Chinese government holds the power

---

[38] Cyberspace Administration of China (4 January 2022). 互联网信息服务算法推荐管理规定 [Regulations on the Management of Algorithm Recommendations for Internet Information Services]. https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm

[39] See the original government website in Chinese: Cyberspace Administration of China (11 December 2022). 互联网信息服务深度合成管理规定 [Regulation on the Administration of Deep Synthesis Internet Information Services]. https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm. Art. 16, 17.
See also the translated version: China Law Translate (11 December 2022). Provisions on the Administration of Deep Synthesis Internet Information Services. https://www.chinalawtranslate.com/en/deep-synthesis/

[40] See the original government website in Chinese: Cyberspace Administration of China (13 July 2023). 生成式人工智能服务管理暂行办法 [Interim Measures for the Management of Generative Artificial Intelligence Services]. http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm
See also the translated version: China Law Translate (13 July 2023). Interim Measures for the Management of Generative Artificial Intelligence Services. https://www.chinalawtranslate.com/en/generative-ai-interim/

[41] Ibid.

to order any AI systems to be suspended if found to have seriously breached the regulations or relevant laws.[42]

The above regulations demonstrate two sides of China's AI governance approach: pro-innovation and pro-stability. Compared to the EU's risk-based approach which (mainly) requires pre-release assessments (largely industry self-assessments) of AI systems, China's approach seems to allow companies to deploy their systems in the market before possible government interventions. In fact, the finalized texts of China's generative AI regulation are considerably less restrictive than a draft version[43] that had circulated previously[44]. This suggests that China's authorities were receptive to public comments and industry concerns as countries race to develop the most powerful LLMs that support AGI tools and applications. As our panelist Professor Zhang observed, despite being the first country to issue a regulation on generative AI, China is very unlikely to implement it in a strict way that might constrain its own technology champions. This pro-innovation approach is understandable given that the government views AI as a key driver for innovation, economic growth, and global competitiveness, which is evident in China's 14th Five-Year Plan (2021-2025) that emphasises AI and technological innovation as key to national development.[45]

At the same time, the Chinese government has been concerned about the potential of digital technologies including AI in undermining social stability. This concern is present in almost all Chinese regulations relevant to internet information services, online platforms, and AI technologies. For example, China's generative AI regulation stipulates that

---

[42] Ibid.

[43] Daum, J. (2023). Key changes to generative AI measures. *China Law Translate*. https://www.chinalawtranslate.com/en/key-changes-to-generative-ai-measures/

[44] Arcesati, R., & Brussee, V. (7 August 2023). China's Censors Back Down on Generative AI. *The Diplomat*. https://thediplomat.com/2023/08/chinas-censors-back-down-on-generative-ai/

[45] State Council of China (13 March 2021). *中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要* [The 14th Five-Year Plan for the National Economic and Social Development of the People's Republic of China and the Outline of Long-Term Goals for 2035] https://www.gov.cn/xinwen/2021-03/13/content_5592681.htm
See the translated version: Asian Development Bank (June 2021). The 14th Five-Year Plan for National Economic and Social Development of the People's Republic of China. https://www.adb.org/publications/14th-five-year-plan-high-quality-development-prc

providing and using generative AI services should not "generate content that may incite subverting state power and socialist systems...and undermine national unity and social stability".[46] In addition, most Chinese regulations on digital technologies, including the above-mentioned regulations relevant to AI technology, require service providers to conduct "security assessments" if their services "have the potential to influence public opinion and mobilize society".[47] As Bradford observes, the Chinese state leverages technology to fuel the country's economic growth and exert control in the name of social stability, both of which are central to the survival of the Chinese Communist Party.[48]

3.2 Commonalities in AI governance values and principles

As discussed above, the major AI powers are adopting approaches to AI governance with different emphases. However, as our panelist Professor Robert Trager, Co-Director of the Oxford Martin AI Governance Initiative at Oxford University, noted, when it comes to principles of AI governance, there are also many "commonalities" among countries and regions. While the "democratic West" and "authoritarian China" are often characterised as having very different values and espousing different principles in AI governance, we suggest that such differences are often exaggerated and commonalities overlooked.

For example, human-centric values are one of the commonalities that are not limited to the EU (and wider western democracies). In China, "putting people first" or "people-centered" (以人为本 yi ren wei ben) is a slogan conceived by the Chinese party-state and proliferated in the political sphere since 2004, meaning development should serve the interests of humans (or the people).[49] In recent years, the Chinese government has frequently used this people-centered rhetoric to describe its approach for AI governance,

---

[46] See for example as in footnote 40, article 4.
[47] See for example, ibid., article 17.
[48] Bradford, A. (2020). *Digital Empires: The Global Battle to Regulate Technology*. Oxford, UK: Oxford University Press. p.24.
[49] Leung, T. T. F., & Tam, C. H. L. (2015). The "Person-Centred" Rhetoric in Socialist China. *The British Journal of Social Work, 45*(5), 1489–1507. http://www.jstor.org/stable/43687926

especially when targeting an international audience. For example, China's Global AI Governance Initiative (GAIGI) states that: [50]

> *We should uphold a people-centered approach in developing AI, with the goal of increasing the wellbeing of humanity and on the premise of ensuring social security and respecting the rights and interests of humanity, so that AI always develops in a way that is beneficial to human civilization. We should actively support the role of AI in promoting sustainable development and tackling global challenges such as climate change and biodiversity conservation.*

What China calls "people-centered" AI may not differ substantially from the EU and US's emphasis on "human-centric AI", both emphasizing that AI should serve the wellbeing of humanity, respect human rights, and promote sustainable development.

Apart from human-centered values, major AI powers (and most other countries) also share many AI principles such as inclusive growth, accountability, transparency, explainability, robustness, security, and safety. These are the principles listed by the OECD, comprising 38 advanced economies.[51] The OECD AI principles so far have been adopted and endorsed by a wide range of countries and international organizations including the US, the EU and the G20 (of which China is a member). Research on Chinese large-scale pre-trained models from 2020 to 2022 found that, among 26 sampled papers, 12 discussed ethics or governance issues, emphasising concerns related to bias and fairness, misuse, environmental harms, and utilizing access restrictions.[52] As Zeng Yi, a member of the United Nations AI Advisory Body and Professor at the Institute of Automation of the Chinese Academy of Sciences, said in a media interview, some Western experts and media organizations often emphasize the differences between

---

[50] Ministry of Foreign Affairs of PRC (20 October 2023). Global AI Governance Initiative. https://www.mfa.gov.cn/eng/wjdt_665385/2649_665393/202310/t20231020_11164834.html
[51] OECD. (nd). AI Principles overview. https://oecd.ai/en/ai-principles
[52] Ding, J. & Xiao, J. (28 April 2023). Recent Trends in China's Large Language Model Landscape. *Centre for the Governance of AI*. https://cdn.governance.ai/Trends_in_Chinas_LLMs.pdf

China and the West in their approaches to AI governance, "but if you really read their ethics and principles, you will realize that commonalities far surpass differences."[53]

Of course, it must be noted that there are still some major differences between the West and China when it comes to values and principles of AI governance. As our panelist Dr Gry Hasselbalch noted, the EU, the US and other "like-minded" partners share *democratic values* in AI governance such as checks and balances on power, a feature that is not shared by authoritarian countries. Another difference is that when emphasizing human rights in AI governance, the EU and the US typically refer to the rights of individual citizens, while in China's context, individual human rights are often discussed together with social stability or collective rights. In addition, China places an emphasis on respecting "national sovereignty" in AI development and governance. As China's GAIGI states, "We oppose using AI technologies for the purposes of…intervening in other countries' internal affairs, social systems and social order, as well as jeopardizing the sovereignty of other states".[54] Here, China's "national sovereignty" principle differs from the EU's "digital sovereignty" strategy, with the former emphasizing non-interference in other nations' internal affairs and the latter emphasizing Europe's ability to act independently in the digital world (i.e., not relying on companies from the United States).

3.3 Values-based governance vs. hardware-based governance

Given that countries and regions share many values and principles of AI governance, is a values-based approach viable in the international domain? This was one of the key debates between our panelists. For Dr Gry Hasselbalch, if countries want to find commonalities on AI governance, then shared democratic values between like-minded partners have to be the starting point. In other words, an approach based on human rights and democratic values is essential for any international framework of AI governance. She noted that there are already many international initiatives based on shared values such

---

[53] Wang, X. (11 July 2020). 国家新一代 AI 治理专委会委员曾毅：中外治理"和而不同" [Member of the National New Generation AI Governance Special Committee Zeng Yi: "Harmony but Difference" in Domestic and Foreign Governance] (Translated from Chinese). *The Paper*. https://www.thepaper.cn/newsDetail_forward_8216294

[54] Ministry of Foreign Affairs of PRC (20 October 2023). Global AI Governance Initiative. https://www.mfa.gov.cn/eng/wjdt_665385/2649_665393/202310/t20231020_11164834.html

as these, including the G7 AI Principles and Code of Conduct, Global Partnership on Artificial Intelligence (GPAI), etc. Through the framework of the EU-US Trade and Technology Council, the EU and the US have agreed on 65 common terms on AI, including "human-centric AI". In her view, these outcomes demonstrate the effectiveness of a values-based approach in facilitating international cooperation around AI governance. Kate Jones, CEO of the Digital Regulation Cooperation Forum of the UK, has a similar view. She has highlighted that human rights law should be the foundation of AI governance to ensure that AI will benefit everyone. [55]

Such a values-based approach was rejected by our panelist Professor Milton Mueller, who emphasized that trying to inject values into global digital infrastructure was exactly the source of conflicts in digital global governance. It is true that the US, the EU, and China have different understandings regarding concepts like privacy and well-being,[56] which explains why some surveillance technologies like face recognition are widely deployed in China (for public safety or identity verification)[57] but are associated with great concern in the EU. Even the US and the EU, often referring to each other as "like-minded partners", could not reach a deal until July 2023 on the EU-US Data Privacy Framework (DPF) after the European Court of Justice invalidated the previous cross border data transfer arrangement in 2020.[58] A previous OXGS report pointed out that the Western approach of trying to incorporate democratic values into global technology standards can lead to a bifurcated system of technology standards, which are intended to be a public good to ensure technology interoperability.[59] However, it is worth noting that there are

---

[55] Jones, K. (10 January 2023). AI governance and human rights: Resetting the relationship. *Chatham House*. https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights

[56] Šucha, V., & Gammel, J. (2021). Humans and Societies in the Age of Artificial Intelligence. European Commission publications. https://data.europa.eu/doi/10.2766/61164, p. 15.

[57] China has begun the process of regulating the use of face recognition technologies. In 2023, China launched a consultation to solicit public opinions on draft regulations specifically targeting facial recognition technology. See reference: Ministry of Justice of PRC (10 August 2023). China mulls first nationwide comprehensive guidelines for use of facial recognition technology. http://en.moj.gov.cn/2023-08/10/c_909702.htm

[58] European Commission (2023). Commercial sector: adequacy decision on the EU-US Data Privacy Framework. https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/eu-us-data-transfers_en

[59] Milne, C., & Wang, J. (2022). *The Geopolitics of Global Technology Standards: Key Issues and Solutions*. Policy Report, Oxford Global Society. http://oxgs.org/wp-content/uploads/2022/11/Report-The-Geopolitics-of-Global-Technology-Standards-Key-Issues-and-Solutions.pdf

increasing efforts to incorporate fundamental human rights within global AI standards which have expanded beyond the traditional scope of technical standards (see section 5.3 for further discussion).[60]

In contrast to this values-based governance mode, some experts are suggesting a hardware-based approach to governing AI (referencing Figure 2.1, DiploFoundation's multilayer governance model). A preprint paper, [61] a collaboration between 19 researchers including academics at leading universities, civil society, and industry players like OpenAI, argues that governing computing power (or "*compute*") could be an effective way to achieve AI policy goals such as safety. This is because computing power is detectable, excludable, and quantifiable. The paper adds that despite the fact that much fundamental research is still needed to validate this approach, governments are increasingly using computing power as a "lever" to pursue AI policy goals, such as limiting misuse risks, supporting domestic industries, or engaging in geopolitical competition. One example is seen in the EU AI Act, which includes a critirion in identifying systemic risks of foundation models based on how much computing power is used to train the model.[62] Our panelist Professor Angela Zhang pointed to the use of semiconductor chips as "leverage" by the US government in geopolitical competition with China in AI (see section 4.2).

4. The role of geopolitics in AI governance

AI is expected to become a crucial component of economic and military power in the future.[63] Not only being a major factor shaping nations' strategies for AI development and governance, geopolitics is also leading to intensified export controls and retaliations,

---

[60] Lewkowciz, G., & Sarf, R. (February 2024). Taking Technical Standardization of Fundamental Rights Seriously for Trustworthy Artificial Intelligence. La Revue des Juristes de Sciences Po n. 25.

[61] Sastry, G., et al, (13 February 2024). *Computing power and the Governance of Artificial Intelligence*. https://doi.org/10.48550/arXiv.2402.08797

[62] Wanat, Z., Gallardo, C., & Smith, T. (7 December 2023). EU to put extra guardrails on AI foundation models like GPT-4. *Sifted*. https://sifted.eu/articles/foundation-model-eu-ai-act

[63] Maslej, N., et al. (April 2023). The AI Index 2023 Annual Report. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA*. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

potentially leading to fragmented AI global governance frameworks, and strengthening asymmetry in power relations between technology haves and have-nots.

4.1 AI geopolitics: beyond the US-China two-way narrative

According to Stanford University's AI Index Report 2023 report, the US and China are the world's top two AI powers, leading in many areas of AI. Below are figures compiled by the report for AI private investment, AI research (based on publication numbers and citations), the number of AI start-ups, and the development or deployment of AI tools and applications:[64]

- In 2022, private AI investment in the US totaled US$ 47.4 billion and was roughly 3.5 times the amount invested by the next highest country, China ($13.4 billion).
- The US leads in terms of the total number of newly funded AI companies, seeing 1.9 times more than the EU and the UK combined, and 3.4 times more than China.
- China leads in total AI publications (accounting for 40% of the world's total AI publications in 2021) and AI journal citations (accounting for 29% in 2010-2021), while the US was ahead in terms of AI conference and repository citations.
- In 2022, 54% researchers working on large language and multimodal models were affiliated with the US (22% from the UK, 8% from China, 6.25% from Canada, 5.8% from Israel).
- China leads in the deployment of some AI tools, such as industrial robot installations. In 2021, China installed more industrial robots than the rest of the world combined.

This is why the US and China are described as the world's two AI superpowers, as predicted by Kaifu Lee, a world-renowned AI expert, in his book *AI Superpowers* published in 2019.[65] Both the US and China have ambitions for AI leadership. This is epitomized in the words of Arati Prabhakar, Director of the White House Office of Science and Technology Policy, who recently said that "we're in a moment where *American*

---

[64] Ibid.
[65] Lee, K. (2019). *AI Superpowers: China, Silicon Valley, and the New World Order*. New York, NY: Harper Business.

leadership in the world depends on American leadership in AI."[66] In 2017, China set itself the goal of becoming a "world-leading" country in AI theory, technology, and applications by 2030.[67] In March 2024, China announced in the Government Work Report[68] that the country would launch the "AI Plus" initiative to promote the integration of AI and the real economy and build digital industry clusters with international competitiveness.[69]

The ambitions of the US and China in AI leadership also manifest in their efforts to influence global AI governance frameworks. The US has been active in multilateral AI initiatives via frameworks like the OECD, G7, GPAI, and the US-EU TTC, while China announced its Global AI Governance Initiative in October 2023 during the third Belt and Road (BRI) International Summit.[70] Our panelist Kayla Blomquist, Director of the Oxford China Policy Lab and doctoral researcher at Oxford Internet Institute, highlighted the deep-rooted tensions among major powers and the formation of "like-minded coalitions" in global AI governance. She observed that, as a result, there are competing stakeholder alliances for AI governance: one led by the West and the other by China through the frameworks of BRI and BRICS Plus.

We deem that it is necessary to go beyond this depiction of US-China two-way competition or rivalry narrative in discussing the dynamics of AI geopolitics (a narrative that was also reflected in the OXGS seminar). First, there is still a substantial distance for China to cover if it is to keep up with the US in AI technology, and some argue that the US is currently the only AI superpower. Many Chinese AI experts acknowledge that China lags the US in areas including foundation models like GPT-4, top AI talent, advanced AI

[66] Murgia, M. (25 January 2024). White House science chief signals US-China co-operation on AI safety. *Financial Times*. https://www.ft.com/content/94b9878b-9412-4dbc-83ba-aac2baadafd9

[67] Carter, W. A., & Crumpler, W. D. (2019). China's National Strategy for AI. In Smart Money on Chinese Advances in AI. *Center for Strategic and International Studies (CSIS)*. http://www.jstor.org/stable/resrep22599.7, p.2

[68] The State Council of PRC (12 March 2024). 政府工作报告 [Government Work Report]. https://www.gov.cn/yaowen/liebiao/202403/content_6939153.htm
See also the translated document: China Daily (13 March 2024). China releases full text of government work report. https://www.chinadaily.com.cn/a/202403/13/WS65f0818aa31082fc043bc400.html

[70] Ministry of Foreign Affairs of PRC (18 October 2023). Foreign Ministry Spokesperson's Remarks on the Global AI Governance Initiative. https://www.fmprc.gov.cn/eng/xwfw_665399/s2510_665401/2535_665405/202310/t20231018_11162874.html

chips, and start-up investment.[71] Of course, China has made rapid progress in some of these areas. For example, by January 2024, there were already over 40 Chinese LLMs approved by the Chinese government regulator for public use, some of which are among the world's leading open-source LLMs in terms of performance.[72]

Second, outside the US and China, so-called "middle powers", such as the UK, Canada, France, Singapore, India, South Korea, and Israel, among others, have become important players with geopolitical aspirations in both AI development and global AI governance. For example, the UK's *National AI Strategy*, published in 2021, states that "The UK is a global superpower in AI and is well placed to lead the world over the next decade as a genuine research and innovation powerhouse".[73] Figure 4.1 illustrates AI research and development links among superpowers and some middle powers. India, having relatively limited foundational capacity for AI infrastructure, recently approved building three semiconductor plants with investments of over US$15 billion, aiming to become a major global chip hub.[74] According to the Government AI Readiness Index 2023, which measures the readiness of governments to integrate AI into public services, many "middle powers" including Singapore, the UK, South Korea, and Finland ranked higher than China.[75] Most middle powers are also investing in AI initiatives/strategies as part of their

---

[71] See for example: Qu, Z. (2 March 2024). *中美 AI 差距在拉大？听听从业者怎么说*
[Has the gap between China and the United States in AI expanded? Let's listen to what the experts say] (Translated from Chinese). *China Business Journal*. https://new.qq.com/rain/a/20240302A00ERZ00
Another example: Huang, L., & Zhang, L. (8 March 2024). *正视中美差距，聚焦通用大模型"主战场"，全国人大代表刘庆峰给出 9 点系统性发展建议* [Facing the gap between China and the United States, focusing on the 'main battlefield' of general large models, National People's Congress deputy Liu Qingfeng offers nine systematic development suggestions] (Translated from Chinese). *The Paper*.
https://www.thepaper.cn/newsDetail_forward_26605212
[72] C114. (29 January 2024). *中国 6 个月来批准逾 40 个人工智能模型，可供公共使用* [China approved over 40 AI models for public use] (Translated from Chinese). https://www.c114.com.cn/ai/5339/a1253993.html
[73] Office of Artificial Intelligence, Government of the United Kingdom (September 2021). National AI Strategy. https://www.gov.uk/government/publications/national-ai-strategy
[74] Press Information Bureau, Gov of India (Feb 2024). Giant leap for India Semiconductor Mission: Cabinet approves three more semiconductor units. https://pib.gov.in/PressReleseDetail.aspx?PRID=2010132
See also the report by Bloomberg: Phartiyal, S. (29 February 2024). India Okays $15 Billion of Milestone Chip Plant Investments. *Bloomberg*. https://www.bloomberg.com/news/articles/2024-02-29/india-approves-15-billion-in-milestone-chip-plant-investments?embedded-checkout=true
[75] Hankins, E., et al. (2023). Government AI Readiness Index 2023. *Oxford Insights*. https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-1.pdf

attempts to shape AI governance. This is evidenced by the first AI Safety Summit held in the UK, a landmark effort to initiate an international response to the safety of frontier AI. Following this summit, South Korea and France have announced plans to host the second (virtual) and third AI Safety Summit in May and November of 2024, respectively, demonstrating their aspirations to play a more central role in global AI governance.
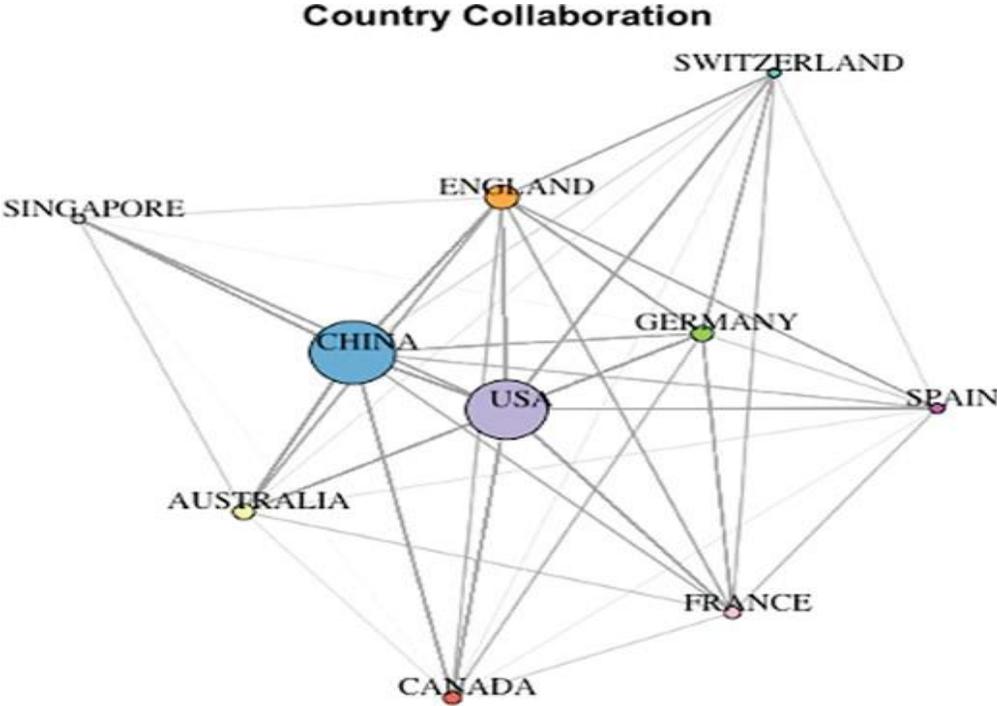
Figure 4.1 Source: Figure 53.1, The Oxford Handbook of AI Governance, Justin B. Bullock (ed.) et al. (19 May 2022) https://doi.org/10.1093/oxfordhb/9780197579329.013.53

4.2 Computing power: a "point of leverage" in AI competition?

Despite the US enjoying a significant advantage in AI, China's fast-moving effort to catch-up is treated with suspicion and fear in Washington, where the government has promulgated a series of measures to deny China's access to advanced AI technology, especially advanced semiconductor chips. On 7 October 2022, the US Department of Commerce's Bureau of Industry and Security (BIS) issued a new list of export controls on

"advanced computing and semiconductor manufacturing items" to China.[76] At the same time, the US has taken measures to incentivise domestic research and the manufacturing of semiconductors in the US in an effort to maintain its technological leadership and ensure the security of the semiconductor supply chain. The CHIPS and Science Act (2022)[77] provides US$52.7 billion for American semiconductor research, development, manufacturing, and workforce development.[78]

The US's control over China's access to advanced AI technology has even expanded to cover semiconductor chips with moderate capabilities[79] and cloud services. According to a recently proposed regulation from the BIS, US cloud service providers will need to report foreign users (mainly Chinese companies) if they tap their computing power for training large AI models.[80] Meanwhile, the US protection of its its own semiconductor industry is extending to "legacy chips" (i.e., current-generation and mature-node semiconductors). In January 2024, the BIS launched a new survey of the US semiconductor supply chain, targeting the use and sourcing of China-manufactured legacy chips.[81] A previous BIS report claims that China's subsidy on chip production constitutes a national security threat to the US.[82] As a response to US trade measures of this kind, China in July 2023

---

[76] Bureau of Industry and Security, US Department of Commerce (October 7, 2022). Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC). https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file

[77] United States Congress (1 July 2022). H.R.4346 - Chips and Science Act. https://www.congress.gov/bill/117th-congress/house-bill/4346

[78] The White House (9 August 2022). FACT SHEET: CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China. https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/

[79] This is seen in the US Commerce Department's intervention on Nvidia's designing of less powerful chips for the Chinese market. See for example: Kan, M. (4 December 2023). Commerce Dept. to Nvidia: Stop Redesigning AI Chips to Sell Them to China. *PCMag UK*. https://uk.pcmag.com/processors/149984/commerce-dept-to-nvidia-stop-redesigning-ai-chips-to-sell-them-to-china

[80] US Department of Commerce (29 January 2024). Taking Additional Steps To Address the National Emergency With Respect to Significant Malicious Cyber Enabled Activities. https://aboutbgov.com/bct3

[81] Bureau of Industry and Security, US Department of Commerce (January 2024). Mature Node Semiconductor Assessment. https://www.bis.doc.gov/index.php/legacy-chips-survey

[82] Bureau of Industry and Security, US Department of Commerce (17 October 2023). Commerce Strengthens Restrictions on Advanced Computing Semiconductors, Semiconductor Manufacturing Equipment, and Supercomputing Items to Countries of Concern. https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3355-2023-10-17-bis-press-release-acs-and-sme-rules-final-js/file

announced export controls on the rare minerals gallium and germanium – both necessary components for manufacturing chips.[83]

For our panelist Professor Mueller, the "main source" of the US-China technological conflict is America's belief that it needs to stop China's participation in the digital ecosystem as a peer competitor and that the competitive threat has been exacerbated by China's retaliatory protectionist measures. He argued that geopolitical tensions are by-products of these states' reactionary attempts to gain power advantage via "digital neo-mercantilism", which is leading to the "weaponisation of interdependence". Also, our panelist Dr Stephen Pattison, Chair at the Internet of Things Security Foundation and former Vice President at Arm, noted that the US' threat perception of China and fear of lagging in the technological race has encouraged it to draw a "very high fence" around its industrial policy. Our panelist Wendell Wallach, Fellow at Carnegie Council for Ethics in International Affairs, observed that if American policies driven by fear of China and China's protectionism continue to confront each other, then the ratcheting up of US-China competition in AI research and markets is inevitable.

It is unsurprising that the US is resorting to computing power as a lever to control China's AI development. AI capabilities are built on three main inputs: computing power (based on semiconductor chips), training data and algorithms. Compared to data and algorithms, computing power is perceived as being easier to control. As Saif Khan and Alexander Mann (2020) point out, "because general-purpose AI software, datasets, and algorithms are not effective targets for controls, the attention [of the U.S. government] naturally falls on the computer hardware necessary to implement modern AI systems."[84] The high concentration of the supply chain of semiconductor chips also makes it convenient for the US to exert such control since some steps in the chips supply chain rely on a very small

---

[83] Ministry of Commerce of PRC (3 July 2023). *商务部 海关总署公告 2023 年第 23 号 关于对镓、锗相关物项实施出口管制的公告* [Ministry of Commerce and General Administration of Customs Announcement No. 23 of 2023 on the Implementation of Export Controls on Gallium and Germanium Related Items] (Translated from Chinese). http://www.mofcom.gov.cn/article/zwgk/gkzcfb/202307/20230703419666.shtml
See also report by Reuters: Home, A. (10 July 2023). China flexes critical metals muscles with export curbs. *Reuters.* https://www.reuters.com/markets/commodities/china-flexes-critical-metals-muscles-with-export-curbs-2023-07-10/
[84] Khan, S.M. (April 2020). AI Chips: What They Are and Why They Matter. *Center for Security and Emerging Technology.* https://cset.georgetown.edu/publication/ai-chips-what-they-are-and-why-they-matter/

number of companies.[85] That is, a single company, the Taiwan-based TSMC, produces the majority of the world's most advanced chips; TSMC, in turn, relies on extreme ultraviolet (EUV) lithography machines that are also only manufactured by the Dutch company ASML; the American company NVIDIA has a market share of over 90% on data center GPU design; and cloud services are dominated by a few large providers including Amazon Web Services, Microsoft Azure and Google Cloud.

The effectiveness of using computing power as "leverage" to contain China's AI capability can, however, be questioned. Citing his previous working experience at Arm, a leading British semiconductor company, our panelist, Dr Pattison, noted that Chinese engineers once came to Arm to learn certain skills, but that is not the case any more. He argued that there is little doubt that China will move towards producing its own advanced AI chips. It is reported that the Chinese semiconductor company SMIC can now produce chips of 7nm (as used in the Kirin 9000s for Huawei Mate 60 Pro).[86] Therefore, it may be that denying China's access to advanced chips will at best lead to a temporary slowing of Chinese progress in AI. In fact, the Biden Administration received domestic criticism for its policy as some revenue that formerly flowed to the US chip companies has been redirected to the Chinese chip-manufacturing companies, which may allow China to double down on its efforts to develop advanced AI chips.[87] Thus, it is not clear how long the US can use computing power as "leverage".

4.3 The concentration of power and divide in AI governance

Apart from US-China rivalry, another major geopolitical concern regarding AI governance is the worrisome concentration of power in the hands of a few countries and a limited number of private companies. This is partly because the threshold for building foundation

---

[85] See Preprint: Heim, L., et al. (14 February 2024). Computing Power and the Governance of Artificial Intelligence. *Centre for the Governance of AI*. https://www.governance.ai/post/computing-power-and-the-governance-of-ai

[86] Allen, G. C. (6 October 2023). In Chip Race, China Gives Huawei the Steering Wheel: Huawei's New Smartphone and the Future of Semiconductor Export Controls. *Center for Strategic and International Studies (CSIS)*. https://www.csis.org/analysis/chip-race-china-gives-huawei-steering-wheel-huaweis-new-smartphone-and-future

[87] Allen, G. C. (11 October 2022). "Choking off China's Access to the Future of AI," *Center for Strategic and International Studies (CSIS)*. https://www.csis.org/analysis/choking-chinas-access-future-ai

AI models that underpin AI applications like ChatGPT is very high, and only a small number of companies that are based in a few countries (predominantly the US but also China) can afford the computing power, have access to huge amounts of data, and are developing complex algorithms to train the models. This high threshold makes it difficult for new entrants to compete. It is reported that OpenAI CEO Sam Altman has been in talks with investors to raise as much as US$5 to $7 trillion for AI chip manufacturing, which is much greater than the entire semiconductor industry's current $527 billion in global sales combined.[88]

The concentration of power in AI is exacerbating the divide between technology haves and have-nots in both AI development and governance. On the one hand, most of the developers in the Global South dealing with AI and related technologies are dependent on big technology firms based in the West. Historically, technologies developed by and for developed countries are often not suitable to be used in lower income countries which have their own baseline of concerns, needs, and social inequalities.[89] Despite this divide, many Global South countries are showing a strong interest in harnessing the potential of AI to enhance their socioeconomic development. For instance, according to a report on the state of AI regulation in Africa, of the 55 African countries, 5 have adopted specific national AI strategies, 15 have established an AI task force (or expert body, agency, or committee), and 6 plan or already have attempted to enact AI laws.[90] It is encouraging to see a South African startup recently launched a multilingual generative AI platform, Vambo AI[91], that supports communication, learning, and content creation in various African languages.

[88] Hagey, K., & Fitch, A. (11 December 2023). Sam Altman Seeks Trillions of Dollars to Reshape Business of Chips and AI. OpenAI chief pursues investors including the U.A.E. for a project possibly requiring up to $7 trillion. *Wall Street Journal*. https://www.wsj.com/tech/ai/sam-altman-seeks-trillions-of-dollars-to-reshape-business-of-chips-and-ai-89ab3db0

[89] Acemoglu, D. (24 April 2020). How the Other Half Automates. *Project Syndicate*. https://www.project-syndicate.org/magazine/artificial-intelligence-in-developing-countries-by-daron-acemoglu-2020-04

[90] Tsebee, D., & Oloyede, R. (March 2024). State of AI Regulation in AI: Trends and Developments. *Tech Hive Advisory and Centre for Law & Innovation.* https://assets-global.website-files.com/641a2c1dcea0041f8d407596/65f7fe8b1f81f94e0dc7643e_State%20of%20AI%20Regulation%20in%20Africa%20Trends%20and%20Developments%20.pdf, p.11.

[91] Indigenous Language AI. https://www.vambo.ai/

However, the lack of AI capabilities also limits Global South countries' participation in global AI governance. So far, the debate on risks and opportunities associated with AI has been largely concentrated in North America and Western Europe.[92] According to the World Economic Forum, AI policy and governance frameworks are predominantly made in Europe, the US and China (46%), compared to only 5.7% in Latin America and 2.4% in Africa.[93] Our panelist Raquel Jorge Ricart, Policy Advisor at Elcano Royal Institute, highlighted this gap, saying that countries from Latin America, the Caribbean, and Africa have not usually been invited to high-level forums on AI governance. As our panelist Dr Hasselbalch noted, one of the policy goals of AI regulation is to address the concentration of power in AI (especially the power concentration in Silicon Valley).

5. Global AI governance: Key issues

Along with rapid progress in AI, there are rising concerns around the world about the risks associated with this transformative technology, be it disinformation, bias, loss of work, or so-called "existential risks". Since many risks, especially the claimed existential ones, are inherently global in nature, they require a global response. Considering the distinctive governance approaches in the US, the EU, and China and the geopolitical tensions around AI, this raises a series of questions: Is a centralized global governance framework possible or necessary? If possible, on what areas should a centralized framework focus? For any global AI governance framework, where should the institutional power lie and what forms should it take?

---

[92] Muggah, R., Seiler, G., & Laforge, G. (2 March 2023). AI and the Global South. *Project Syndicate*. https://www.project-syndicate.org/commentary/ai-governance-first-principles-must-include-global-south-by-robert-muggah-et-al-2023-03

[93] World Economic Forum (January 2024). Generative AI Governance: Shaping a Collective Global Future. Briefing Paper Series, p. 11. https://www3.weforum.org/docs/WEF_AI_Governance_Alliance_Briefing_Paper_Series_2024.pdf

5.1 A centralized or distributed governance framework?

As our panelist Professor Robert Trager put it, the first question to ask regarding a global AI governance framework is: *what areas of AI should be placed under international governance*? In his view, national autonomy should be preserved in many areas of AI governance, as different regulatory approaches reflect distinctive emphases on values (see section 3.1). But in other areas such as LAWS[94], national regulation cannot be effective without international governance. Thus, he argued that on the international level, societies need to identify what areas need international governance and where international consensus can be reached. In response to Trager's question, our panelist Dr Stephen Pattison proposed differentiating AI systems based on the levels of risk they pose: "existential" risks (e.g., military systems like the LAWS) and lower risks (e.g., most civilian systems). He argued that for AI systems that involve existential risks, a centralized global framework is needed; while for most civilian AI systems, where values and cultural norms may differ in emphasis, they may best be dealt with at a national or regional level.

Dr Pattison's differentiated approach to global AI governance aligned with the views of some panelists who supported a distributed governance framework for most civilian AI systems. As discussed previously, different regulatory approaches and distinctive emphases on underlying values are factors behind conflicts between countries. For this reason, our panelist Dr Gry Hasselbalch, who advocated a values-based governance approach, argued that parallel international governance frameworks on AI are an "asset", rather than an obstacle. Another reason, as Professor Trager noted, is that countries are at different stages of AI development and thus have different national interests, with some countries hesitant to introduce any regulations that may inhibit innovation. In this regard, the principle stated in China's GIAIG – "All countries, regardless of their size, strength, or social system, should have equal rights to develop and use AI"[95] – aligns well with the interest of Global South countries. This may also explain why the Association of

---

[94] United Nations (2023). Lethal Autonomous Weapon Systems (LAWS). https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/
[95] Ministry of Foreign Affairs of PRC (20 October 2023). Global AI Governance Initiative. https://www.mfa.gov.cn/eng/wjdt_665385/2649_665393/202310/t20231020_11164834.html

Southeast Asian Nations (ASEAN) has adopted a voluntary and business-friendly approach towards AI governance[96], despite the EU's lobbying for other parts of the world to align with its own stricter proposed framework.[97] For our panelist Raquel Jorge Ricart, a distributed global governance framework would also be beneficial for diversity since a centralized framework could reduce the voices from different countries and stakeholders. She suggested that international AI development initiatives should prioritize issues such as the inclusion of languages from vulnerable and smaller communities.

Regarding a centralized or distributed framework on AI governance, our panelist Wendell Wallach provided a nuanced argument. As he noted, a distributed AI governance framework is already the reality, "whether we like it or not". However, this distributed governance may result in hundreds of different stakeholder groups. There will therefore be a need for an observatory that convenes coordinated discussion that helps people understand the state of AI and flags issues that must be addressed in the coming years. We agree that some form of centralized global governance on lower risk AI systems is needed to coordinate between national and multilateral governance frameworks. Given that countries and regions share many values and principles of AI governance such as transparency and non-discrimination, such a framework seems feasible. Take the governance of algorithms and end-user protection (the upper two levels of governance, see Figure 2.1) as an example. As discussed in section 3.1, both China and the EU require online platforms or other providers of algorithmic recommendation systems to allow users to opt-out from personalized recommendations (user autonomy), to disclose their recommendation mechanisms or parameters (algorithmic transparency), and not to discriminate against users based on certain characteristics (user protection). They also require AI systems to make users aware when AI technology is being used (AI transparency). These common (or similar) rules could form the foundation of a global governance framework for AI algorithms and user protection. A unitary governance framework is especially important for industry players. As our panelist Maxime Ricard,

---

[96] Association of Southeast Asian Nations (February 2024). ASEAN Guide on AI Governance and Ethics. https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf
[97] Hutt, D. (2024). AI regulations: What can the EU learn from Asia? *DW*. https://www.dw.com/en/ai-regulations-what-can-the-eu-learn-from-asia/a-68203709

Policy Manager at Allied for Startups (representing the interests of startups and SMEs in the EU and worldwide), observed, businesses, especially start-ups, do not like to navigate fragmented governance frameworks and "want to be global and scalable from day one".

5.2 Can countries cooperate on existential risks?

Existential risks associated with AI have become a more prominent topic in recent years in the wake of technology breakthroughs like ChatGPT and intelligent robots. An open letter from the Centre for AI Safety (CAIS) states that "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemic and nuclear war."[98] Signatories of this statement include many AI scientists from leading AI labs like OpenAI and Google DeepMind, research institutions like the Alan Turing Institute, and public figures like Bill Gates, while there has also been criticism of their view of the likelihood of a risk of extinction of human beings.

While a global governance framework for AI systems with existential risks may be needed, our panelists debated whether such a framework is achievable in the foreseeable future. As discussed, one major obstacle to cooperation in dealing with perceived existential risks from AI is geopolitical rivalry between the West and China. As our panelist Professor Angela Zhang pointed out, before the first AI Safety Summit in November 2023, China had been excluded from most important international discussions on AI, except those hosted by the UN. In fact, China's participation in the summit was a topic that generated much controversy and criticism in the UK. Some UK lawmakers questioned whether China – "a technological rival and military threat" to the West – should be included in meetings about sensitive technology.[99] Professor Zhang stressed that without "having China at the table" and without the US and China setting out to cooperate, international cooperation on many aspects of AI governance is "going nowhere". The importance of

---

[98] Hinton, G. et al. (30 May 2023). Statement on AI Risk: Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war. *Center for AI Safety*. https://www.safe.ai/work/statement-on-ai-risk#open-letter

[99] Goh, B., & Sandle, P. (4 November 2023). Exclusive: China took part in leaders' AI meeting even though UK did not acknowledge. *Reuters*. https://www.reuters.com/technology/china-took-part-leaders-ai-meeting-even-though-uk-did-not-acknowledge-2023-11-03/

"having China at the table" resonated with other panelists. Wendell Wallach considered China's participation in AI governance as "essential", saying that "wooing" China should be central to any global effort for AI governance. Dr. Pattison stated that [the West] designing a system that excludes China is "synonymous with undermining its effectiveness from the start".

Despite some pessimism around the possibility of China-US cooperation on AI governance regarding any AI system, as voiced by our panelist Professor Mueller and some others[100], there are also some signs of movement towards cooperation. The passing of the Bletchley Declaration during the first AI Safety Summit, of which China was among the signatories, is one such sign. The Declaration affirmed the urgency of deepening the understanding of and actions to address the substantial risks arising from "frontier AI models".[101] Our panelist Professor Trager considered the result of this summit as encouraging, and he was optimistic about the potential for China and the US to sit down to discuss issues involving certain risks, just as the Soviet Union and the US once cooperated on the reduction of strategic nuclear weapons[102]. In addition, China seems to be keen to participate in global discussion on risks of AI that may be designated as "existential". As a report from Concordia AI (a leading AI safety research social enterprise in China) notes, China is assembling top Computer Science scholars to evaluate the existential risks of AI. Currently, numerous Chinese labs are conducting research on AI safety and expert discussions around frontier AI risks have become more mainstream in China since 2022.[103] In 2023, multiple Chinese experts, including those from AI labs like SenseTime and ByteDance and research institutions like Tsinghua University and Chinese Academy of Sciences, signed the Future of Life Institute (FLI) and Center for AI

[100] See for example: Huq, A. (11 March 2024). A world divided over Artificial Intelligence. *Foreign Affairs*. https://www.foreignaffairs.com/united-states/world-divided-over-artificial-intelligence

[101] UK Government (1-2 November 2023). The Bletchley Declaration by Countries Attending the AI Safety Summit. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[102] Arms Control Association (2022). *U.S.-Russian Nuclear Arms Control Agreements at a Glance*. https://www.armscontrol.org/factsheets/USRussiaNuclearAgreements#:~:text=The%20Strategic%20Arms%20Reduction%20Treaty,warheads%20as%20counted%20using%20the

[103] Concordia AI (October 2023). State of AI Safety in China. https://concordia-ai.com/wp-content/uploads/2023/10/State-of-AI-Safety-in-China.pdf

Safety (CAIS) open letters on the risks of frontier AI. [104] During their meeting in November 2023, American President Joe Biden and Chinese President Xi Jinping affirmed the need to address the risks of advanced AI systems through US-China government talks.[105]

Some are more cautious about the future of global cooperation on dealing with AI existential risks. One area that needs particular attention is the LAWS. Given the military potential of AI, geopolitical competition may exacerbate the rise of autonomous weapons, and many worry that deep-fake technologies may mean that national leaders make decisions based on false information. Countries such as the US and China are exploring AI's potential in warfare. According to a report on the military application of AI, the market for autonomous weapons is projected to be around a US $26.36 billion by 2027.[106] Some experts consider the introduction of autonomous weapons as the third revolution in warfare, after gunpowder and nuclear weapons.[107] A report, commissioned by the US State Department (but which does not represent the views of the US government), concludes that the U.S. government must move "quickly and decisively" to avert an existential-level threat from AI. However, instead of advising the US government to cooperate with others to deal with this threat, the report recommends that the government should outlaw powerful open-source AI models and further tighten controls on the manufacture and export of AI chips to control the spreading of advanced AI technology.[108] These measures, if adopted, would only intensify geopolitical conflicts around AI. Our panelist Dr Pattison commented that one lesson the world needs to learn is how to act together to deal with AI risks *before* a disaster, rather than only after a disaster happens as with chemical and nuclear weapons. Our panelist Wendell Wallach argued that only if

[104] Hinton, G. et al. (30 May 2023). Statement on AI Risk: Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war. *Center for AI Safety*. https://www.safe.ai/work/statement-on-ai-risk#open-letter

[105] White House (15 November 2023). Readout of President Joe Biden's Meeting with President Xi Jinping of the People's Republic of China. https://www.whitehouse.gov/briefing-room/statements-releases/2023/11/15/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-2/

[106] Verified Market Research (September 2020). Global military artificial intelligence (AI) market size and forecast. https://www.verifiedmarketresearch.com/product/military-artificial-intelligence-ai-market/

[107] Li, K. (11 September 2021). The Third Revolution in Warfare. *The Atlantic*. https://www.theatlantic.com/technology/archive/2021/09/i-weapons-are-third-revolution-warfare/620013/

[108] Perrigo, B. (11 March 2024). Exclusive: U.S. Must Move 'Decisively' to Avert 'Extinction-Level' Threat From AI, Government-Commissioned Report Says. *Time.* https://time.com/6898967/ai-extinction-national-security-risks-report/

the threats are real enough in the minds of the policymakers can they bring China, the US and other countries together on AI governance.

5.3 Where should institutional power lie?

As our panelist Professor Robert Trager pointed out, there are a lot of discussions around AI global governance and many similar principles have been proposed, but so far not many results have been produced. Thus, the question should be how to move "*from principle to practice*". We agree with Professor Trager that it is time to explore how to implement shared AI principles (as discussed in 3.2), exploring issues such as the institutional arrangements and the forms of implementation to ensure accountability.

Is the UN the right institution for global AI governance?

The UN's *The Interim Report: Governing AI for Humanity* argues for a global governance framework led by the UN due to its unique legitimacy. It states that the UN is "uniquely positioned" to address the distinctively global challenges and opportunities presented by AI because it is a body with universal membership founded on the UN Charter and has a commitment to embracing the diversity of all peoples of the world; thus it can turn "a patchwork of evolving initiatives into a coherent, interoperable whole".[109] This argument certainly has some truth to it and is supported by many countries of the Global South, and China has stressed that "an international AI governance institution should be set up under the UN framework and all countries should be able to participate on equal terms".[110]

However, whether the UN should hold the institutional power for convening the governace of AI is an issue of debate. Our panelist Wendell Wallach thought that UN participation adds to the legitimacy of any global governance initiative, but AI governance led by the UN alone may not be that effective. Our panelist Dr Stephen Pattison, a former senior British diplomat who spent years on UN-related work, was skeptical about a UN-led

---

[109] UN Secretary-General's high-level AI advisory Body (December 2023). Interim Report: Governing AI for Humanity, pp. 1-4.
https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf
[110] Ministry of Foreign Affairs of PRC (7 March 2024). Wang Yi on Global AI Governance: Ensure that AI is a Force for Good, Ensure Safety and Ensure Fairness.
https://www.fmprc.gov.cn/eng/zxxx_662805/202403/t20240308_11256430.html

framework. As he put it, there is a danger that discussions within the UN could be "embroiled in political negotiations". For instance, some countries might be wary that governance efforts are being plotted to stop them getting access to AI technologies. He commented that some issues such as existential risks of AI are "too important to give it to the UN where it might get embroiled in committees for too long". A Brookings commentary, published in February 2024, argues that, while the UN has a crucial role to play in global AI governance, any approach to global governance must be "distributed and iterative", involving participation by a wide range of stakeholders including governments and *private companies* (UN consists of member states).[111]

In addition to the UN, other institutional arrangements and mechanisms such as those of the International Atomic Energy Agency (IAEA) were suggested during the OXGS seminar. Our panelist Professor Trager deemed that the IAEA may be a useful model for governing AI systems like the LAWS. He noted that one advantage of this mechanism is that the responsible authority does not need to go into all countries, but only some countries to inspect whether they are developing LAWS. Thus, it could be an effective way (and less intrusive to most countries) to control the risks associated with dangerous autonomous weapons.

Should we go for hard or soft governance?

Another important aspect in implementing global AI governance is the form of implementation: hard (mandatory) or soft (voluntary) governance. As discussed earlier, except for perceived existential risks, hard regulation may not be popular with many countries, especially those with scarce AI capabilities. Soft AI governance includes voluntary Codes of Conduct, guidelines, and global AI standards. Currently, many international organizations such as the IEEE, ISO, IEC, and WHO are developing their standards on AI technologies, applications, and management, with the ISO/IEC 42001 standard on AI management systems (AIMS)[112] being a particularly relevant example.

---

[111] Kerry C. F., et al. (2024). Should the UN Govern Global AI? *Brookings*.
https://www.brookings.edu/articles/should-the-un-govern-global-ai/
[112] ISO (December 2023). ISO/IEC 42001:2023. Information technology. Artificial intelligence. Management system https://www.iso.org/standard/81230.html

Although there is much competition within and among these organizations, there is certainly some hope for global cooperation. American official Arati Prabhakar recently commented that, while China and the US may disagree on certain values and approaches to regulation, "there will also be places where we can agree", including global technical and safety standards for AI software.[113]

However, as some of our panelists argued, soft governance must be backed by some forms of enforcement to ensure the accountability of AI governance. Professor Trager pointed to international governance mechanisms like the International Civil Aviation Organization (ICAO) and the Financial Action Task Force (FATF) as useful references for AI governance. As he noted, when there is a violation of ICAO standards somewhere, one practice is that the Federal Aviation Administration (FAA) in the US can deny access to US airspace for any flight originating from a country that is in violation of ICAO standards. The FATF, which monitors global money laundering and terrorist financing trends, maintains black and gray lists that can lead to sanctions on and substantial reputational harm to non-compliant nations.[114] Professor Trager suggested that, if a jurisdiction violates international standards, then other jurisdictions could refuse to trade AI products with it. Similarly, our panelist Stephen Pattison suggested that commercial alliances jointly representing complete value chains for AI supply could admit members only on agreed conditions, whose violation could bar access to the latest AI advances. Our panelist Wendell Wallach also supported a hybrid model of governance (both soft and hard) to increase the potential effectiveness of AI governance.

In fact, voluntary AI standards are increasingly incorporated into binding laws and regulations. For example, the EU's AI Act requires providers of high-risk AI systems to conduct risk assessments, and it allows certain companies to demonstrate compliance with this obligation if they voluntarily adhere to European harmonized standards on risk

---

[113] Murgia, M. (25 January 2024). White House science chief signals US-China co-operation on AI safety. *Financial Times*. https://www.ft.com/content/94b9878b-9412-4dbc-83ba-aac2baadafd9

[114] Financial Action Task Force. (nd). Black and grey lists. https://www.fatf-gafi.org/en/countries/black-and-grey-lists.html

management after they are published.[115] As an article from Carnegie Endowment for International Peace notes, requirements in the EU AI Act, like other EU legislations, are outlined at a high level, leaving significant room for standards to fill in the blanks in implementation.[116] Since AI risk assessment involves ethical issues like bias and discrimination, it means that harmonized AI standards will be used to attest to compliance in the areas of ethics and human rights; traditionally, technical standards are mainly used to ensure product safety or interoperability. The impact and effectiveness of this new approach have yet to be seen.

Conclusions

Given distinctive values and governing priorities, it is unsurprising for major AI powers (and other countries/regions) to adopt regulatory approaches with diffrent emphases towards AI governance. For instance, a central goal of the EU's AI governance is to protect individuals' rights and safety; while in China, the protection of the rights and interests of individual users and businesses always comes together with an emphasis on social stability, with the latter often being prioritized. For the US, the goal of facilitating innovation and maintaining its technology leadership is consistently prioritised in its strategies and policies on AI development and governance. Despite these differences, we emphasize the commonalities between countries/regions in the values and principles of AI governance which are often overlooked. While China's AI governance approach reflects its authoritarian state, it shares many values with its Western counterparts including the human-centered values. These commonalities could be the foundation for global collaboration on AI governance.

The debate around values-based and hardware-based governance modes reflects the tensions in AI governance. While values-based AI governance can facilitate multilateral cooperation, it also leads to the confrontation of different "like-minded" coalitions or

---

[115] European Commission (24 January 2024). AI Act: Consolidated Text. https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-Overview_24-01-2024.pdf

[116] Pouget, H., & Zuhdi, R. (2024). AI and Product Safety Standards Under the EU AI Act. *Carnegie*. https://carnegieendowment.org/2024/03/05/ai-and-product-safety-standards-under-eu-ai-act-pub-91870

alliances, which makes global AI governance more challenging. In contrast, while hardware-based AI governance, due to its distinct features (detectable, quantifiable, and excludable), can be a useful tool to reduce misuse and address safety concerns in both domestic and global governance, it can also be deployed to contain other countries' AI development. This will incentivize further geopolitical rivalry and could provoke retaliation, leading to further disruption of global AI supply chains.

A single or centralized global governance framework should be developed to deal with existential/extreme risks, bringing together different stakeholder groups. We deem it possible to establish such a framework, because countries share human-centered values and many positive signs — the passing of the Bletchley Declaration and the signing of open letters regarding AI existential risks by influential scientists, researcher institutions and leading AI labs around the world — have been witnessed. Lesser levels of risk may better be handled via a variety of regional and national approaches, in a continuation and expansion of existing initiatives. However, this does not mean no global framework is needed to address lower risks, and some global mechanisms will be useful to establish common norms and standards, helping businesses to navigate global markets.

Given its universal membership and commitment to diversity, we agree that the UN should play a crucial role in global AI governance, which is also beneficial to address the extreme concentration of power in AI. However, for existential risks of AI, inspiration for institutional arrangements and mechanisms can be drawn from existing international governance models such as the IAEA and ICAO. These mechanisms may prove to be more effective in dealing with urgent risks, as they are less enmeshed in political negations than the UN. As for the form of global governance, we deem that global AI standards and other soft laws will have important roles to play. However, a hybrid governance model—soft laws backed by binding laws and regulations—will be most effective to ensure the accountability of AI governance.